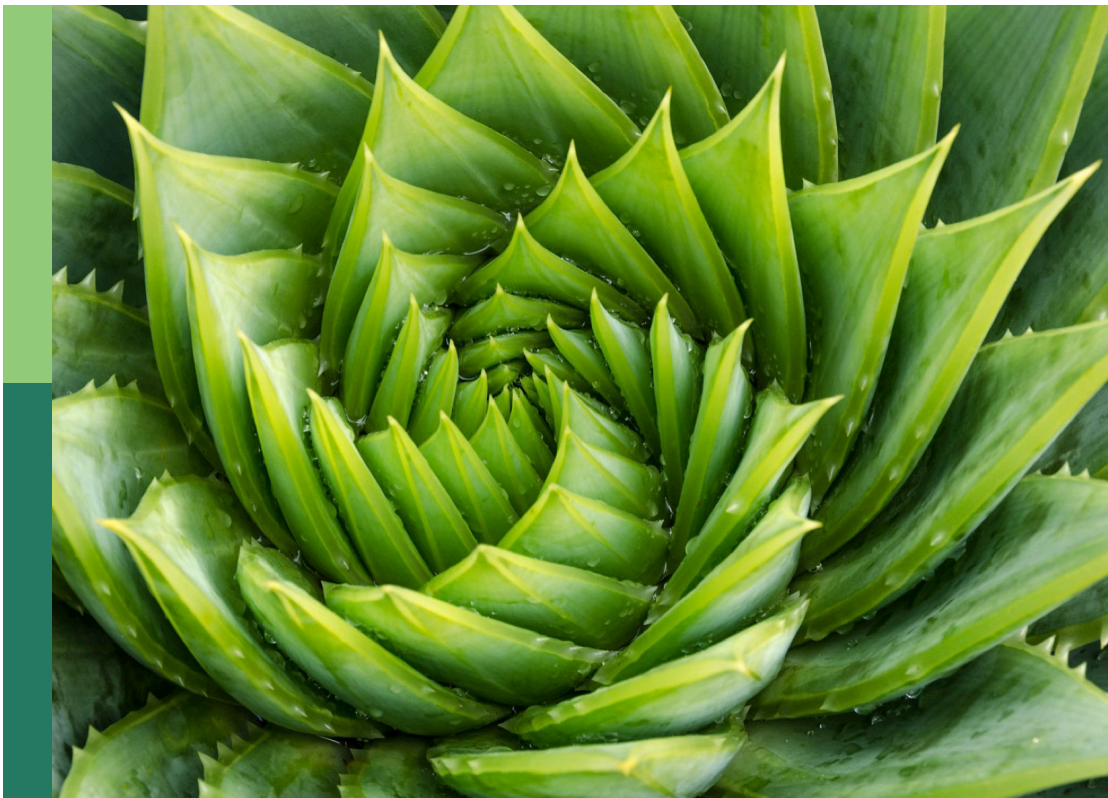
frontiers | Research Topics

Genetic characterization of yield- and quality-related traits in legumes

Edited by
Qijiian Song, Jianghua Chen, Zhengjun Xia,
Kyuya Harada and Chuanen Zhou

Published in
Frontiers in Plant Science



ISSN 1664-8714
ISBN 978-2-8325-3828-9
DOI 10.3389/978-2-8325-3828-9

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

Genetic characterization of yield- and quality-related traits in legumes

Topic editors

Qijiian Song — Agricultural Research Service, United States Department of Agriculture, United States

Jianghua Chen — Key Laboratory of Tropical Plant Resource and Sustainable Use, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences (CAS), China

Zhengjun Xia — Chinese Academy of Sciences (CAS), China

Kyuya Harada — Osaka University, Japan

Chuanen Zhou — Shandong University, China

Citation

Song, Q., Chen, J., Xia, Z., Harada, K., Zhou, C., eds. (2023). *Genetic characterization of yield- and quality-related traits in legumes*. Lausanne: Frontiers Media SA.

doi: 10.3389/978-2-8325-3828-9

Table of contents

05 Editorial: Genetic characterization of yield- and quality-related traits in legumes
Zhengjun Xia, Qijian Song, Kyuya Harada, Jianghua Chen and Chuanen Zhou

09 Using transcriptomic and metabolomic data to investigate the molecular mechanisms that determine protein and oil contents during seed development in soybean
Wenjing Xu, Qiong Wang, Wei Zhang, Hongmei Zhang, Xiaoqing Liu, Qingxin Song, Yuelin Zhu, Xiaoyan Cui, Xin Chen and Huatao Chen

22 Genome-wide association mapping of seed oligosaccharides in chickpea
Dinakaran Elango, Wanyan Wang, Mahender Thudi, Sheelamary Sebastiar, Bharathi Raja Ramadoss and Rajeev K. Varshney

35 Genome-wide characterization of *AINTEGUMENTA-LIKE* family in *Medicago truncatula* reveals the significant roles of *AINTEGUMENTAs* in leaf growth
Xiao Wang, Juanjuan Zhang, Jing Zhang, Chuanen Zhou and Lu Han

47 GmTOC1b inhibits nodulation by repressing *GmNIN2a* and *GmENOD40-1* in soybean
Yuhang Zhang, Qun Cheng, Chunmei Liao, Lanxin Li, Chuanjie Gou, Zheng Chen, Yanan Wang, Baohui Liu, Fanjiang Kong and Liyu Chen

59 *MINI BODY1*, encoding a MATE/DTX family transporter, affects plant architecture in mungbean (*Vigna radiata* L.)
Xin Li, Yahui Jia, Mingzhu Sun, Zikun Ji, Hui Zhang, Dan Qiu, Qiao Cai, Yan Xia, Xingxing Yuan, Xin Chen and Zhenguo Shen

72 Identification of major QTLs for soybean seed size and seed weight traits using a RIL population in different environments
Shilin Luo, Jia Jia, Riqian Liu, Ruqian Wei, Zhibin Guo, Zhandong Cai, Bo Chen, Fuwei Liang, Qiuju Xia, Hai Nian and Yanbo Cheng

89 Fine mapping and candidate gene analysis of proportion of four-seed pods by soybean CSSLs
Fubin Cao, Ruru Wei, Jianguo Xie, Lilong Hou, Chaorui Kang, Tianyu Zhao, Chengcheng Sun, Mingliang Yang, Ying Zhao, Candong Li, Nannan Wang, Xiaoxia Wu, Chunyan Liu, Hongwei Jiang and Qingshan Chen

104 Morpho-biochemical characterization of a RIL population for seed parameters and identification of candidate genes regulating seed size trait in lentil (*Lens culinaris* Medik.)
Haragopal Dutta, Shivaprasad K. M., Muraleedhar S. Aski, Gyan P. Mishra, Subodh Kumar Sinha, Dunna Vijay, Manjunath Prasad C. T., Shouvik Das, Prashant Anupama-Mohan Pawar, Dwijesh C. Mishra, Amit Kumar Singh, Atul Kumar, Kuldeep Tripathi, Ranjeet Ranjan Kumar, Sanjeev Gupta, Shiv Kumar and Harsh Kumar Dikshit

118 Genome-wide characterization of aldehyde dehydrogenase gene family members in groundnut (*Arachis hypogaea*) and the analysis under saline-alkali stress
Xiaoming Zhang, Jingwen Zhong, Liang Cao, Chunyuan Ren, Gaobo Yu, Yanhua Gu, Jingwen Ruan, Siqi Zhao, Lei Wang, Haishun Ru, Lili Cheng, Qi Wang and Yuxian Zhang

129 Developmental dynamic transcriptome and systematic analysis reveal the major genes underlying isoflavone accumulation in soybean
Heng Chen, Changkai Liu, Yansheng Li, Xue Wang, Xiangwen Pan, Feifei Wang and Qiuying Zhang

145 Divergence of functions and expression patterns of soybean bZIP transcription factors
Lin Yue, Xinxin Pei, Fanjiang Kong, Lin Zhao and Xiaoya Lin

157 Genome-wide identification of the soybean cytokinin oxidase/dehydrogenase gene family and its diverse roles in response to multiple abiotic stress
Yanli Du, Zhaoning Zhang, Yanhua Gu, Weijia Li, Weiyu Wang, Xiankai Yuan, Yuxian Zhang, Ming Yuan, Jidao Du and Qiang Zhao

172 Genome-wide association analysis reveals the optimal genomic regions for pod size in bean
Mao Li, Xinyi Wu, Baogen Wang, Xiaohua Wu, Ying Wang, Jian Wang, Junyang Dong, Jian Wu, Zhongfu Lu, Yuyan Sun, Wenqi Dong, Jing Yang and Guojing Li

frontiers | Frontiers in Plant Science

TYPE Editorial
PUBLISHED 17 October 2023
DOI 10.3389/fpls.2023.1281138

OPEN ACCESS

EDITED AND REVIEWED BY
Shaojun Dai,
Shanghai Normal University, China

*CORRESPONDENCE
Zhengjun Xia
xiazhj@iga.ac.cn

RECEIVED 22 August 2023
ACCEPTED 10 October 2023
PUBLISHED 17 October 2023

CITATION
Xia Z, Song Q, Harada K, Chen J and Zhou C (2023) Editorial: Genetic characterization of yield- and quality-related traits in legumes.
Front. Plant Sci. 14:1281138.
doi: 10.3389/fpls.2023.1281138

Editorial: Genetic characterization of yield- and quality-related traits in legumes

Zhengjun Xia[1]*, Qijian Song[2], Kyuya Harada[3], Jianghua Chen[4] and Chuanen Zhou[5]

[1]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, [2]United States Department of Agriculture- Agriculture Research Service (USDA ARS), Soybean Genome & Improvement Lab, Beltsville, MD, United States, [3]Department of Biotechnology, Graduate School of Engineering, Osaka University, Suita, Japan, [4]CAS Key Laboratory of Tropical Plant Resources and Sustainable Use, CAS Center for Excellence in Molecular Plant Sciences, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Kunming, China, [5]The Key Laboratory of Plant Development and Environmental Adaptation Biology, Ministry of Education, School of Life Sciences, Shandong University, Qingdao, China

KEYWORDS

legume, yield, quality, soybean, gene cloning, functional characterization, Medicago, plant architecture

Editorial on the Research Topic
Genetic characterization of yield- and quality-related traits in legumes

Legumes (Fabaceae) are essential crop plants throughout the world, second in importance only to members of the grass family (Poaceae) (Smýkal et al., 2015). One-third of human dietary protein comes from grain legumes, which represent more than one-fourth of global crop production (Smýkal et al., 2015). Legumes are vital for agriculture, natural ecosystems, and agroforestry owing to their capacity for symbiotic nitrogen fixation (Smýkal et al., 2015). The complex nature of traits associated with yield and quality poses challenges for studying their underlying molecular mechanisms. However, gene cloning and functional studies of traits that are unstable and dependent on the environment are being enhanced through various omics technologies. There were a total of 26 submissions for this Research Topic on the genetic characterization of yield- and quality-related traits in legumes, 13 of which were eventually accepted for publication. These papers fall into five subtopics: (1) plant architecture and leaf development, (2) pod and seed traits, (3) seed composition, (4) abiotic stress resistance, and (5) nodulation.

1 Plant architecture and leaf development

The plant architecture of members of the legume family starkly contrasts with that in the grass family, Poaceae. Even though it appears that only a few key genes were targeted by selection during plant domestication and breeding, plant architecture can confer optimal plasticity in response to environmental changes. Moreover, plant architecture has considerable effects on the seed yield of grain legumes, a major source of human dietary protein, which is poised to become increasingly crucial as human diets shift toward reduced

animal protein consumption for health and environmental reasons. However, there has been limited research on the molecular regulation of plant architecture in legumes to date.

Li X. et al. investigated the roles of MATE/DTX (Multidrug and toxic compound extrusion/detoxification) family transporters in controlling plant development as well as stress responses. The authors identified a *mini body 1* (*mib1*) mutant in mung bean (*Vigna radiata*) that shows increased branching, five-leaflet compound leaves, and reduced pod length. Through map-based cloning, they found that *MIB1* encodes a MATE/DTX family protein in mung bean. Despite expression in all tissues, quantitative reverse transcription PCR (qRT-PCR) results demonstrated that young inflorescences exhibit the highest expression of this gene. A complementation assay in *Escherichia coli* suggested that MIB1 functions as a MATE/DTX transporter, while the short-pod phenotype of the Arabidopsis (*Arabidopsis thaliana*) *dtx54* mutant could be partially rescued by heterologous expression of mung bean *MIB1*. Transcriptome analysis indicated that the molecular function of MIB1 might involve a phytohormone pathway. Taking their results together, the authors concluded that MIB1 is important for controlling the establishment of plant architecture in mung bean.

Meanwhile, in a study of the AINTEGUMENTA-LIKE (AIL) family in *Medicago truncatula*, a model legume for functional studies of leaf development, Wang et al. revealed that AINTEGUMENTA (ANT) proteins have important roles in leaf growth. Four *MtANT* genes were present among the 11 *MtAIL* genes they identified. High expression levels of *MtANT1* were observed in tissue containing active stem cells, e.g., the shoot apical meristem and leaf primordium. Further study of *MtANT* quadruple mutants as well as plants overexpressing *MtANT* revealed that MtANTs control leaf size by regulating cell proliferation during secondary morphogenesis in *M. truncatula* leaves. This study is the first to reveal the function of MtANTs during leaf growth. Further work will help to elucidate the molecular mechanism whereby MtANTs regulate leaf size. Because biomass is critical for imrpoving forage grass quality, genetic manipulation of these genes could enhance the biomass production of legume forage crops.

2 Pod and seed traits

Seed size and shape are important quality traits of lentil (*Lens culinaris* Medik.), affecting the yield of milled grain, market grade, and cooking time. Pod traits are important components of yield in snap bean (*Phaseolus vulgaris*) and other fresh legume crops. There are rich genetic resources for seed and pod traits such as size and color among different leguminous species and even within a given species, e.g., common bean. Functional studies of the regulatory networks in different leguminous species will shed light on the molecular bases of seed and pod traits.

Luo et al. focused on quantitative trait locus (QTL) mapping and gene mining for two closely correlated seed traits, seed size and seed weight, in soybean (*Glycine max*) grown in different environments. They identified 18 environmentally stable QTLs among 85 QTLs associated with seed size and weight, which they mapped using a recombinant inbred line (RIL) population developed from Guizao1×B13 (GB13). Notably, the newly identified QTL qSL-3-1 showed a stable effect, contributing to 10.0% to 15.91% of phenotypic variance (PV) in seed length. In addition, qSW20-3 had a stable effect on seed width, explaining 9.22% to 21.93% of PV. Functional annotation and GO enrichment analysis identified 15 candidate genes with possible roles in controlling seed size and weight in soybean. These results provide a reference for studying the development of soybean seed, which will influence molecular breeding and enable consistent enhancement of soybean yield.

The proportion of four-seeded pods (PoFSP) is an important factor contributing to soybean yield. The *Ln* gene was previously shown to pleiotropically control leaflet shape and seed number in soybean. Cao et al. performed fine mapping and candidate gene analysis of the PoFSP in soybean using a chromosome segment substitution line (CSSL) population. Eleven QTLs were identified in the CSSL population, with all 14 genes annotated in the delimited QTL intervals showing variation in the promoter region or coding sequence. Five candidate genes were differentially expressed in pooled accessions displaying opposing phenotypic extremes (PoFSP >35.92% for high pool; PoFSP <17.56% for low pool). Haplotype analysis was consistent with this finding. The results of this work will greatly facilitate the study of candidate genes affecting soybean PoFSP and provide a foundation for marker-assisted selection of the four-seeded pod trait.

Limited information is available on the molecular basis of pod dimension or size in common bean (snap bean). Li M. et al. used genome-wide association analysis to identify the optimal genomic regions for improving pod size in snap bean. Analysis of 88 snap bean accessions they revealed 57 SNPs significantly linked with the pod size trait. Of the 26 candidate genes for pod development, which included cytochrome P450 family genes and WRKY and MYB transcription factor genes, eight showed high expression levels in flowers as well as young pods. These findings increase our understanding of the genetic basis of pod size in snap bean.

Dutta et al. identified genes controlling seed size in lentil through morpho-biochemical characterization of a RIL (F5:6) population derived from a cross between L830 (20.9 g per 1000 seeds) and L4602 (42.13 g per 1000 seeds). A parental polymorphism survey using two extreme phenotype pools and 394 simple sequence repeats (SSRs) identified 31 polymorphic primers. The marker PBALC449 could clearly differentiate the parents from the pool of small seed size. A few candidate genes were identified in the PBALC449 anchored region that might play roles in regulating seed size, including genes encoding ubiquitin C-terminal hydrolase and E3 ubiquitin ligase, as well as other proteins or enzymes such as hexosyltransferase and TIFY-like protein. This study thus improves our understanding of a genomic region controlling seed size in lentil.

3 Seed composition

Understanding the molecular mechanisms that regulate seed composition in legumes is important not only for improving seed

quality but also for enhancing the nutritional and physiological value of these crops.

Xu et al. investigated the molecular regulation of protein and oil contents in soybean seed via transcriptomic and metabolomic analyses. The authors identified 12,712 differentially expressed genes and 315 differentially accumulated metabolites during seed development in two soybean cultivars with different protein and oil contents. KEGG enrichment analysis indicated that the enriched pathways included plant hormone signal transduction, phenylpropanoid biosynthesis, linoleic acid metabolism, glycerolipid metabolism, carbon metabolism, and the biosynthesis of amino acids and secondary metabolites. The authors proposed that soybean varieties with high seed protein contents are prone to delayed leaf senescence as well as earlier photomorphogenesis.

Raffinose family oligosaccharides (RFOs) are abundant in seeds and are difficult for humans and animals to digest, causing flatulence and severe abdominal discomfort. Most leguminous seeds, such as chickpea (*Cicer arietinum*), have noticeable amounts of RFOs. Elango et al. used high-performance liquid chromatography to measure the contents of RFOs (raffinose and stachyose), ciceritol, and sucrose in chickpea and identified genes involved in sugar metabolism and transport via genome-wide association mapping. Accessions having lower RFOs and higher sucrose concentrations might be desirable for use in breeding.

Much is known about the genetic factors regulating isoflavone concentrations in soybean seeds; these bioactive compounds are important for both plants and humans. Chen et al. analyzed the dynamic variation in isoflavone contents and the isoflavone accumulation patterns in soybean at the physiological level using eight RILs. They also investigated the whole-genome expression profiles of four lines as well as 42 meta-transcriptome datasets and identified molecular modules strongly associated with isoflavone concentration and genes affecting isoflavone accumulation in developing seeds. Their findings increase our understanding of the biosynthesis and molecular control of isoflavones at the physiological and molecular levels and will assist in breeding new soybean cultivars having higher isoflavone concentrations.

4 Abiotic stress resistance

Like soybean seeds, the seeds of the legume crop peanut, or groundnut (*Arachis hypogaea*), are rich in protein and oil. Zhang X et al. characterized the aldehyde dehydrogenase (ALDH, EC: 1.2.1.3) family in *A. hypogaea*. Of the 71 members of the peanut ALDH superfamily (*AhALDH*), some genes exhibited tissue-specific expression, and some were differentially expressed upon exposure to saline-alkali stress, implying their possible involvement in plant responses to abiotic stress. These findings can be utilized to breed peanut with improved stress resistance.

Cytokinin oxidase/dehydrogenases (CKXs) regulate plant growth and development by irreversibly degrading cytokinin, thus helping plants cope with environmental stress. Du et al. characterized the GmCKX gene family in soybean using transcriptome profiling (RNA-seq), qRT-PCR, and bio-informatics tools. Eighteen *GmCKX* genes were identified in soybean and clustered into five clades, some of which exhibited tissue-specific expression. RNA-seq analysis demonstrated that GmCKXs may have crucial functions in plant responses to salt and drought stress. Abiotic stress reduced zeatin contents in soybean radicles but increased CKX activity. External application of the phytohormones 6-benzylaminopurine (6-BA) and indole-3-acetic acid (IAA; auxin) promoted transcriptional abundance of *GmCKX10* and *GmCKX18* but hampered the expression of *GmCKX1*, *GmCKX6*, and *GmCKX9*, reducing the zeatin contents in radicles.

Basic leucine zipper domain (bZIP)–containing transcription factors are key factors in regulating environmental signaling and stress responses, as well as carbon–nitrogen balance. Yue et al. reviewed the roles of 161 bZIP family members in soybean, which could be classified into 13 groups. Physiological analyses and genetic engineering have revealed the functions of several soybean bZIP transcription factors. However, the biochemistry and regulatory mechanisms of most bZIP transcription factors in soybean remain unclear.

5 Nodulation

Symbiotic nitrogen fixation is an important component of the nitrogen cycle. Enhancing this process in legumes could facilitate sustainable agriculture, as symbiotic nitrogen fixation contributes almost 20% of the nitrogen needed for global grain and oilseed production.

Nodulation is controlled via an exquisite network consisting of various nodulation-related genes. Zhang Y et al. characterized the function of the TOC1 family member GmTOC1b during nodulation in soybean. Mutants of *GmTOC1b* displayed more numerous and heavier nodules than the wild type due to increased root hair curling and infection thread production, whereas overexpressing *GmTOC1b* inhibited nodulation. Furthermore, the authors determined that GmTOC1b binds to the promoters of nodulation-related genes such as *GmNIN2a* and *GmENOD40-1* to negatively regulate their transcriptional abundance. These findings reveal the crucial function of GmTOC1b in controlling nodule formation in soybean.

The studies highlighted here illuminate the genetic basis of yield- and quality-related traits in legumes, potentially spurring new research on these exciting topics and facilitating the genetic improvement of many important crops.

Author contributions

ZX: Writing – original draft. QS: Writing – review & editing. KH: Writing – review & editing. JC: Writing – review & editing. CZ: Writing – review & editing.

Funding

This work was supported by the Intergovernmental International Science, Technology, and Innovation Cooperation Key Project (Project No. 2022YFE0130300) of the National Key R&D Program (NKP).

Acknowledgments

We are grateful to all of the authors who submitted their manuscripts on this Research Topic. We also appreciate all reviewers and editors for their time.

Reference

Smýkal, P., Coyne, C. J., Ambrose, M. J., Maxted, N., Schaefer, H., Blair, M. W., et al. (2015). Legume crops phylogeny and genetic diversity for science and breeding. *Crit. Rev. Plant Sci.* 34, 43–104. doi: 10.1080/07352689.2014.897904

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

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 29 September 2022
DOI 10.3389/fpls.2022.1012394

OPEN ACCESS

EDITED BY
Qijian Song,
Agricultural Research Service (USDA), United States

REVIEWED BY
Anuradha Singh,
Department of Plant, Soil and Microbial Sciences, Michigan State University, United States
Enrique Martinez Force,
Institute for Fats (CSIC), Spain

*CORRESPONDENCE
Huatao Chen
cht@jaas.ac.cn

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Functional and Applied Plant Genomics, a section of the journal Frontiers in Plant Science

RECEIVED 05 August 2022
ACCEPTED 13 September 2022
PUBLISHED 29 September 2022

CITATION
Xu W, Wang Q, Zhang W, Zhang H, Liu X, Song Q, Zhu Y, Cui X, Chen X and Chen H (2022) Using transcriptomic and metabolomic data to investigate the molecular mechanisms that determine protein and oil contents during seed development in soybean. *Front. Plant Sci.* 13:1012394. doi: 10.3389/fpls.2022.1012394

Using transcriptomic and metabolomic data to investigate the molecular mechanisms that determine protein and oil contents during seed development in soybean

Wenjing Xu[1,2†], Qiong Wang[1†], Wei Zhang[1], Hongmei Zhang[1], Xiaoqing Liu[1], Qingxin Song[3], Yuelin Zhu[2], Xiaoyan Cui[1], Xin Chen[1] and Huatao Chen[1*]

[1]Institute of Industrial Crops, Jiangsu Academy of Agricultural Sciences, Nanjing, China, [2]College of Horticulture, Nanjing Agricultural University, Nanjing, China, [3]State Key Laboratory of Crop Genetics and Germplasm Enhancement, National Center for Soybean Improvement, Jiangsu Collaborative Innovation Center for Modern Crop Production, Nanjing Agricultural University, Nanjing, China

Soybean [*Glycine max* (L.) Merri.] is one of the most valuable global crops. And vegetable soybean, as a special type of soybean, provides rich nutrition in people's life. In order to investigate the gene expression networks and molecular regulatory mechanisms that regulate soybean seed oil and protein contents during seed development, we performed transcriptomic and metabolomic analyses of soybean seeds during development in two soybean varieties that differ in protein and oil contents. We identified a total of 41,036 genes and 392 metabolites, of which 12,712 DEGs and 315 DAMs were identified. Analysis of KEGG enrichment demonstrated that DEGs were primarily enriched in phenylpropanoid biosynthesis, glycerolipid metabolism, carbon metabolism, plant hormone signal transduction, linoleic acid metabolism, and the biosynthesis of amino acids and secondary metabolites. K-means analysis divided the DEGs into 12 distinct clusters. We identified candidate gene sets that regulate the biosynthesis of protein and oil in soybean seeds, and present potential regulatory patterns that high seed-protein varieties may be more sensitive to desiccation, show earlier photomorphogenesis and delayed leaf senescence, and thus accumulate higher protein contents than high-oil varieties.

KEYWORDS

protein content, oil content, gene expression pattern, metabolic pathway, molecular regulation, soybean

Introduction

Soybean [*Glycine max* (L.) Merri.] is one of the most valuable global crops. Vegetable soybean is a type of soybean typically harvested during the R6-R7 stage, at which its pods and seeds are suitable for eating (Young et al., 2000). Soybean seeds contain approximately 40% protein and 20% oil (Chaudhary et al., 2015; Li et al., 2015). Optimizing carbon flux towards the synthesis of fatty acids (FAs) and amino acids and improving seed quality has always been a major objective of soybean breeding programs (Bhati et al., 2021). However, the negative correlations between seed protein content with seed oil content and seed yield have hindered progress (Chung et al., 2003; Patil et al., 2017).

Soybean seed oil is mostly comprised of triacylglycerols (TAGs), which have three acyl groups stemming from five fatty acids: linolenic acid, stearic acid, linoleic acid, oleic acid, and palmitic acid (Clemente and Cahoon, 2009). In plants, TAG biosynthesis entails *de novo* FA biosynthesis within plastids as well as TAG assembly in the endoplasmic reticulum (ER) (Bates, 2016; Xu and Shanklin, 2016). The enzyme acetyl-CoA carboxylase initiates the *de novo* FA biosynthesis pathway by converting acetyl-CoA to malonyl-CoA (Salie and Thelen, 2016). The newly synthesized FAs are activated by conversion to FA acyl-CoAs by long-chain acyl-CoA synthetase (LACS), and are then transported to the ER for TAG biosynthesis.

The protein of soybean seeds has 18 amino acids and includes all nine essential amino acids. Of these, there is a deficiency of Cys, Trp, and Met (Zhang et al., 2018). These amino acids are essential for seed development metabolism. While free amino acids (FAAs) are involved in synthesizing seed-storage proteins, one of their most important roles, they are also precursors to the secondary metabolite biosynthesis and provide energy. In addition, amino acids are efficiently catabolized through the tricarboxylic acid (TCA) cycle (Less and Galili, 2009; Kirma et al., 2012).

The levels of individual amino acids vary greatly during seed development. In *Arabidopsis*, young seeds accumulate mostly Ser, Glu, Gln, Gly, and Ala at 6 days after flowering (DAF). However, at 11 DAF, Leu and Val levels significantly increase. Subsequently, higher Ser and Gly levels were observed at 16 DAF (Baud et al., 2002).

Amino acids in plants are synthesized *via* branched pathways (Less and Galili, 2008; Pratelli and Pilot, 2014). The carbon skeleton for Gln, Glu, proline (Pro), and Arg biosynthesis comes from the Krebs cycle intermediate 2-oxoglutarate. The first precursor for synthesizing the following six amino acids is oxaloacetate: methionine (Met), threonine (Thr), asparagine (Asp), isoleucine (Ile), lysine (Lys), and aspartate (Asn). Pyruvate is used to synthesize alanine (Ala), valine (Val), and leucine (Leu), while tryptophan (Trp), tyrosine (Tyr), and phenylalanine (Phe), all of which are aromatic amino acids, are derived from the shikimate pathway. A precursor of serine (Ser) is 3-phosphoglycerate, which leads then to synthesis of glycine (Gly) and cysteine (Cys).

Previous studies used a variety of populations and various mapping methods to identify 248 and 327 quantitative trait loci (QTL) for soybean seed protein and seed oil content, respectively (SoyBase, https://www.soybase.org). Genome-wide association studies (GWAS) were used to find novel loci for soybean oil and protein contents (Li et al., 2018; Lee et al., 2019; Zhang et al., 2021). However, the mechanisms underlying soybean seed development and the regulation of protein and oil biosynthesis have not been comprehensively investigated.

In order to obtain a more complete understanding of the genetic basis of seed oil and protein accumulation in soybean, we conducted transcriptomic and metabolomic analyses of developing soybean seeds. We identified potential key regulators that regulate the biosynthesis of protein and oil in soybean seed and present potential regulatory patterns. Our study provides a valuable resource for the genetic improvement of soybean seed quality through molecular breeding.

Materials and methods

Plant materials and tissue preparation

For this experiment, we selected two soybean cultivars, namely, 'NPS233' and 'NPS301', to evaluate the variations in protein and oil contents. 'NPS233' accumulates more protein and less oil in the seeds compared to 'NPS301', which is a high oil/low protein variety (Table 1). The plant materials were cultivated at the Luhe experimental base of Jiangsu Academy of Agricultural Sciences in Nanjing, China, in the summer of 2021. Seed samples were collected at four developmental stages; 7 DAF (days after flowering), 14 DAF, 21 DAF, and 28 DAF, frozen in liquid nitrogen and stored at -80°C for further metabolite determinations and extraction of RNA.

Measurement of protein and oil

For measurement of samples protein and oil content (%), ~20 g seeds were grounded to powder and analyzed by Kjeldahl Method and Soxhlet extraction, respectively (Bremner, 1960; Fehr et al., 1968). Samples protein and oil content (%) were averaged over three replications.

TABLE 1 Phenotypic differences between 'NPS233' and 'NPS301'.

Material	Leaf type	Whole growth period (day)	HSW (g)	Plant height (cm)	Number of main stem nodes	Oil content (μg/mg)				Protein content (μg/mg)			
						7 DAF	14 DAF	21 DAF	28 DAF	7 DAF	14 DAF	21 DAF	28 DAF
NPS233	oval	105	25.8	69	18	69.6	73.0	65.9	81.9	42.3	116.7	151.7	205.6
NPS301	circular	120	20.1	44	12	75.6	94.3	126.5	124.1	33.9	73.5	120.0	169.5

HSW, Hundred-seed weight; DAF, days after flowering.

Extracting of metabolites

Biological samples were vacuum freeze-dried with a lyophilizer (Scientz-100F) and then crushed with a mixer mill (MM 400, Retsch) and a zirconia bead at 30 Hz for 1.5 min. 100 mg samples of lyophilized powder were dissolved in a solution of 1.2 ml 70% methanol, vortexed every 30 minutes for 30 seconds (six total vortexes), and maintained overnight at 4°C. Extracts were filtered (SCAA-104, 0.22μm pore size; ANPEL, Shanghai, China; http://www.anpel.com.cn/) after they were centrifuged for 10 min at 16,000 g. They were then used in UPLC-MS/MS analysis.

UPLC conditions

Samples were analyzed with a UPLC-ESI-MS/MS system (UPLC, Shimadzu Nexera X2, www.shimadzu.com.cn/; MS, Applied Biosystems 4500 Q TRAP, www.appliedbiosystems.com.cn/) using the following conditions: UPLC; the column was an Agilent SB-C18 (1.8 μm, 2.1 mm×100 mm); mobile phase used solvent A, pure water with 0.1% formic acid, and solvent B, and acetonitrile with 0.1% formic acid. Samples were separated using a gradient program starting at 95% A, and 5% B. After 9 min, a linear gradient to 5% A, 95% B was set, and a composition of 5% A, 95% B was maintained for 1 min, after which a composition of 95% A, 5% B was attained after 1.10 min and maintained for 2.9 min. The flow velocity, column oven temperature, and injection volume were 0.35 ml per minute, 40° C, and 4 μl, respectively. An ESI-triple quadrupole-linear ion trap (QTRAP)-MS was connected with the effluent.

ESI-Q TRAP-MS/MS

Triple quadrupole (QQQ) and LIT were obtained using a triple quadrupole-linear ion trap mass spectrometer (Q TRAP), AB4500 Q TRAP UPLC/MS/MS System, with an ESI Turbo Ion-Spray interface that was placed in positive and negative ion modes and operated using Analyst 1.6.3 software (AB Sciex). The parameters of the ESI source operation included the following: ion source, turbo spray; source temperature, 550°C; ion spray voltage, (IS) 5500 V (positive ion mode)/-4500 V (negative ion mode); ion source gas I (GSI), gas II (GSII), and curtain gas (CUR) were set at 50, 60, and 25.0 psi, respectively; collision-activated dissociation (CAD) was high. Instrument tuning and mass calibration were carried out using 10 and 100 μmol/L polypropylene glycol solutions in QQQ and LIT modes, respectively. QQQ scans were obtained with MRM experiments with collision gas (nitrogen) set to medium. DP and CE for individual MRM transitions were performed using additional optimization of DP and CE. Certain MRM transitions were observed during each time period according to the eluted metabolites within that time frame.

Principal component analysis

PCA was not supervised and was carried out with the statistics function prcomp in R (www.r-project.org). Prior to this analysis, the data were subjected to unit variance scaling.

Pearson correlation coefficients and hierarchical cluster analysis

Sample and metabolite HCA results are displayed as heatmaps with dendrograms, and the Pearson correlation coefficients (PCC) of the samples were assessed using the cor function in R and only displayed as heatmaps. The R package pheatmap was used to perform both PCC and HCA. For the HCA, a color spectrum was used to display the normalized metabolite signal intensities with unit variance scaling.

Identifying differential metabolites

Metabolites that were significantly regulated among groups were identified using VIP ≥1 and absolute Log_2FC ($|Log_2FC| \geq 1$). The values of the VIP were obtained based on the results of OPLS-DA, which was produced with the R package MetaboAnalyst R and also provided the score plots and permutation plots. Prior to OPLS-DA, all data were mean-

centered and log transformed (log_2). A permutation test with 200 permutations was conducted to prevent overfitting.

Analysis of enrichment and KEGG annotation

The KEGG Compound database (http://www.kegg.jp/kegg/compound/) was used to annotate the resulting metabolites, which were subsequently mapped to the KEGG Pathway database (http://www.kegg.jp/kegg/pathway.html). Pathways with significantly regulated metabolites were then inputted into MSEA (metabolite sets enrichment analysis), and the p-values from a hypergeometric test were used to assess their significance.

RNA extraction and transcriptome sequencing

Total RNA was extracted from soybean samples using the RNAprep Pure Plant Kit (TSINGKE, TSP412). RNA quantification and cDNA library construction were performed according to the methods described by Lu et al. (2018). The libraries were sequenced using Illumina NovaSeq S6000. Three biological replicates were included in each experiment.

Quality control of data

Fast p v 0.19.3 was used to sort the original data, and filter DNA sequencing reads with adapters. Reads with Ns (unknown base calls) higher than 10% of base read numbers were discarded, as were sequencing reads with >50% low-quality (Q ≤20). The resulting clean reads were used to perform all subsequent analyses.

Mapping reads to reference genomes

Reference genomes and their related annotation files were obtained from the designated website, while HISAT v2.1.0 was used to generate the index. Clean reads were cross-referenced with the reference genome (Wm82.a2.v1).

Quantifying levels of gene expression

FeatureCounts v1.6.2 was used to analyze gene alignment, after which gene length was used to establish the FPKM (fragments per kilobase of transcript per million mapped reads) for each gene, which is the most common method of assessing levels of gene expression.

Analyzing differences

The differential expression between the two groups was analyzed using DESeq2 v1.22.1, while the *P* value was corrected with the Benjamini & Hochberg method. The resulting *P* values and $|log_2$ fold-change$|$ were used to categorize significant differences in gene expression. Differentially expressed genes were assessed according to the following parameters: corrected *P*-value of 0.05 and absolute fold change ≥2 (Mao et al., 2005).

Analysis of differential gene enrichment

The hypergeometric test was used to perform enrichment analysis, while DEG enrichment in KEGG pathways was assessed with KOBAS software (Mortazavi et al., 2008). GO (Gene Ontology) was performed using the GOseq (v3.10.1) (Götz et al., 2008).

Results

Metabolome profiling

To generate comprehensive metabolic regulatory networks of soybean seed at different developmental periods, we collected soybean seed samples from two soybean varieties at 7 days after flowering (DAF), 14 DAF, 21 DAF, and 28 DAF. Seeds of the soybean variety 'NPS233' accumulate more protein and less oil than seeds of 'NPS301' at all four seed developmental periods (Figures 1A, B). The experiment consisted of three biological replicates, each of which was a pool of seeds from five plants.

The metabolomes of the 24 samples were profiled using the widely-targeted UPLC-MS/MS metabolic profiling approach. We detected 392 compounds that could be grouped into 11 classes, including 25 alkaloids, 58 amino acids and derivatives, 52 flavonoids, 68 lipids, 39 nucleotides and derivatives, 32 organic acids, 30 phenolic acids, 23 terpenoids, 16 saccharides and alcohols, 14 vitamins, and 35 other compounds (Table S1). Among the metabolite classes, lipids, amino acids and derivatives, and flavonoids were the most abundant.

We performed hierarchical clustering analysis of the 24 samples (7, 14, 21, and 28 DAF for the two varieties), and the results showed that the three biological replicates from each developmental stage grouped together, which suggested that the generated metabolome data was highly reliable (Figure 1C). The metabolites were clustered into three main groups, indicating that there are distinct accumulation levels among the samples taken at four stages of seed development (7, 14, 21, and 28 DAF).

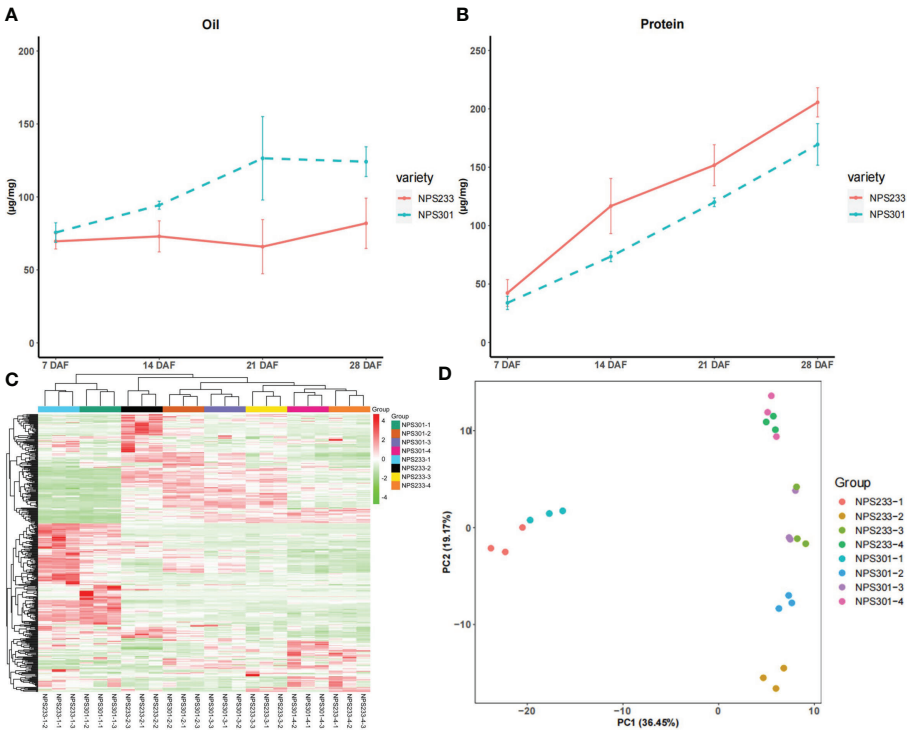


FIGURE 1
Phenotypic difference and 392 detected metabolites between 'NSP233' and 'NPS301'. **(A)** Protein content of 7 DAF, 14 DAF, 21 DAF and 28DAF. **(B)** Oil content of 7 DAF, 14 DAF, 21 DAF and 28 DAF. **(C)** The hierarchical clustering analysis of the 24 soybean seed samples. **(D)** Principal component analysis of the seed samples, each circle represents a sample. For hierarchical clustering analysis, normalized signal intensities of metabolites (unit variance scaling) are visualized as a color spectrum.7 DAF, NPS233-1 and NPS301-1; 14 DAF, NPS233-2 and NPS301-2; 21 DAF, NPS233-3 and NPS301-3; 28 DAF, NPS233-4 and NPS301-4.

The results of principal component analysis (PCA) grouped the eight samples into four discreet clusters (Figure 1D). The first principal component (PC1, 36.45%) effectively separated the 7 DAF samples from the other three samples, implying that there are changes in metabolite accumulation during soybean seed development.

Analysis of the differentially accumulated metabolites

To identify the differentially accumulated metabolites (DAMs) between each pairwise comparison of soybean seed samples, we analyzed three biological replicates of seeds from the two soybean varieties at four different growth periods (NPS233-1, NPS233-2, NPS233-3, and NPS233-4; NPS301-1, NPS301-2, NPS301-3, and NPS301-4). Metabolites with variable importance in projection (VIP) ≥1 and fold-change ≥2 or ≤0.5 were considered to be DAMs. We performed two separate types of analyses of the DAMs; the first between the different growth periods of the same variety, and the second between the two varieties at the same growth period. As a result, 315 DAMs were detected; in the first group of analyses these included 209, 203, and 203, and 179, 179, and 192 DAMs in NPS233-1 vs. NPS233-2, NPS233-1 vs. NPS233-3, and NPS233-1 vs. NPS233-4, and NPS301-1 vs. NPS301-2, NPS301-1 vs. NPS301-3, and NPS301-1 vs. NPS301-4 (Table S2). In the second group of analyses, we detected 69, 104, 45, and 37 DAMs in the NPS301-1 vs. NPS233-1, NPS301-2 vs. NPS233-2, NPS301-3 vs. NPS233-3, and NPS301-4 vs. NPS233-4 comparisons, respectively (Table S2). Among the DAMs, we detected 49 flavonoids, 48 lipids, and 44 amino acids and derivatives, and these were the most abundant.

Comparative analysis of the six groups of DAMs identified 274 and 235 DAMs in the NPS233 and NPS301 comparisons within the groups of seed samples, respectively. Moreover, 69 and 30 metabolites were exclusively differentially accumulated in the NPS233 and NPS301 groups, respectively. There were 20 differentially accumulated lipids and six differentially accumulated amino acids and derivatives in the NPS233 groups; we found that one phospho-sugar and one sugar acid, D-fructose 6-phosphate and D-glucuronic acid, were differentially accumulated exclusively in NPS233. Also, one lipid and five amino acids and derivatives were exclusive to the NPS301 groups, but no differentially-accumulated saccharides were found.

Furthermore, we investigated DAMs with different accumulation patterns, which included some intermediates of the TCA cycle; for example, α-ketoglutaric acid, a ketone derivative of glutaric acid, was up-regulated in NPS233 versus NPS301 in developmental period 2 (14 DAF). α-ketoglutarate is its carboxylate and is also known as 2-oxoglutarate; it is a keto acid generated by deaminating glutamate and an intermediate during the Krebs cycle.

We next performed KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway enrichment analysis. The top enriched KEGG terms annotated for all the comparisons were "flavone and flavonol biosynthesis", "pyrimidine metabolism", "isoflavonoid biosynthesis", "biosynthesis of amino acids", "biosynthesis of secondary metabolites", "ABC transporters", "zeatin biosynthesis", and "aminoacyl-tRNA biosynthesis" (Figure S1).

In addition, DAM comparisons between the samples from NPS301 revealed an enrichment of KEGG terms related to "cyanoamino acid metabolism," "2-oxocarboxylic acid metabolism", and "pyruvate metabolism". The top enriched KEGG terms in the comparisons of the NPS233 samples were "flavonoid biosynthesis", "folate biosynthesis", "galactose metabolism", "biotin metabolism", "alanine, aspartate and glutamate metabolism", and "cysteine and methionine metabolism" (Figure S1).

Differentially expressed gene analysis

To investigate the molecular mechanisms underlying the regulation of protein and oil biosynthesis in the soybean seed developmental periods, we performed RNA-seq analysis for the two varieties. A total of 47.0 million clean DNA sequencing reads were generated, of which 94.79% had a Phred quality score of Q30 or greater (Table S3). The mapping statistics of the sequencing libraries are summarized in Table S4. We identified 1,176 novel genes, of which 394 were successfully annotated. In total, 41,036 genes were found to be expressed in at least one sample (Table S5).

The dataset was used to identify differentially expressed genes (DEGs). We performed two separate analyses of the DEGs; the first was between the different growth periods of the same variety, and the second was between the two varieties at the same growth period. The pairwise comparisons of samples from NPS301 and NPS233 detected 12,906 DEGs in total (Table S6). The pairwise comparisons NPS233-1 vs. NPS233-2, NPS233-1 vs. NPS233-3, and NPS233-1 vs. NPS233-4 identified 3,341 (1,887 upregulated; 1,454 downregulated), 4,015 (2,314 upregulated; 1,701 downregulated) and 3,991 (2,228 upregulated; 1,763 downregulated) DEGs, respectively. The pairwise comparisons NPS301-1 vs. NPS301-2, NPS301-1 vs. NPS301-3, and NPS301-1 vs. NPS301-4 identified 7,889 (4,539 upregulated; 3,350 downregulated), 7,730 (4,505 upregulated; 3,225 downregulated) and 8,234 (4,747 upregulated; 3,487 downregulated) DEGs, respectively (Figure 2A; Table S6).

We identified 1,676 DEGs across the three compared groups from NPS233 and 4,420 DEGs from the comparative analysis of the NPS301 samples (Figures 2B, C), suggesting that these core conserved DEGs may be associated with the accumulation of soybean protein and oil contents, respectively.

We detected 655 DEGs through pairwise comparisons between the two soybean varieties at the same growth period, including 242, 366, 236, and 169 DEGs in NPS301-1 vs. NPS233-1, NPS301-2 vs. NPS233-2, NPS301-3 vs. NPS233-3, and NPS301-4 vs. NPS233-4, respectively (Table S6).

To further analyze the potential function of these DEGs, we conducted a KEGG pathway enrichment analysis. The top enriched KEGG terms annotated for all the compared groups were "carbon metabolism", "glycerolipid metabolism", "linoleic acid metabolism", "plant hormone signal transduction", "biosynthesis of secondary metabolites", "glycosylphosphatidylinositol (GPI)-anchor biosynthesis", "beta-alanine metabolism", and "other glycan degradation".

Notably, the DEGs in the three NPS233 comparison groups were enriched in the KEGG terms "biosynthesis of amino acids", "ribosome", "glycolysis/gluconeogenesis", "nitrogen metabolism", "fatty acid degradation", and "pantothenate and CoA biosynthesis". KEGG pathway enrichment analysis of DEGs in the pairwise comparisons of NPS301 samples indicated that these genes were involved in several metabolic processes including "fatty acid elongation", "sphingolipid metabolism", "fatty acid metabolism", "pyruvate metabolism", and "citrate cycle (TCA cycle)" (Figure S2).

DEGs that respond to soybean seed development

To gain further insights into gene expression changes that occurred over the four soybean seed developmental periods, we performed k-means clustering of the DEGs based on their expression patterns and obtained 12 clusters (Figure S3; Table S7).

The DEGs in clusters 1, 6, 7, and 8 showed similar expression trends in which the expression levels gradually decreased after the second development period (samples NPS301-2 and NPS233-2). DEGs in clusters 3 and 12 showed an increase in expression in the first three periods, after which the expression of these DEGs decreased in the fourth period. Notably, the DEGs in clusters 3 and 12 showed different expression levels in NPS233 and NPS301, indicating that the DEGs in cluster 3 may be mainly related to protein content, while the DEGs in cluster 12 may be mainly related to oil content. The relative expression levels of DEGs in clusters 2, 4, 9, 10, and 11 were down-regulated over the four developmental periods. Among them, the DEGs in cluster 11 showed a rapid

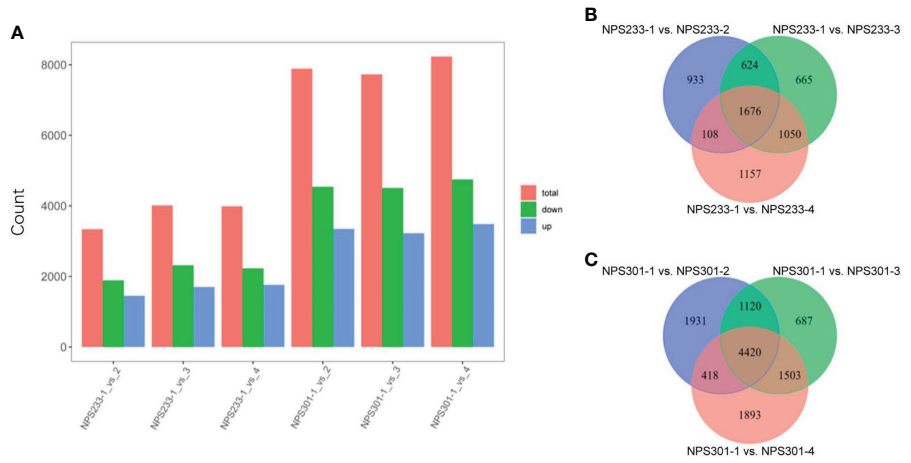


FIGURE 2
Differentially expressed genes in the six compared groups from NPS233 and NPS301. **(A)** Number of DEGs identified through the comparative analysis. **(B, C)** Venn diagrams depicting the DEGs between pairwise comparisons of seed samples from NPS233 and NPS301.

down-regulation of expression, and the DEGs in clusters 2 and 4 had opposite expression patterns in NPS233 and NPS301, indicating that these DEGs are strongly associated with oil or protein contents in soybean seeds. Finally, expression levels of DEGs in cluster 5 were rapidly up-regulated, which indicated that these genes are involved in soybean seed development.

These clusters were selected as candidate DEG sets, and four candidate DEG sets containing 2,887 DEGs were further analyzed. We performed a KEGG pathway enrichment analysis of these genes. Enrichment analysis of the DEGs in cluster 3 indicated that they were mainly enriched in protein processing in the endoplasmic reticulum, biosynthesis of the amino acids valine, leucine, and isoleucine, protein export, galactose metabolism, glycosaminoglycan degradation, and other glycan degradation (Figures 3A, B), implying that the DEGs in cluster 3 may be mainly related to protein biosynthesis. *Glyma.06G169700* in cluster 3, which was functionally annotated as encoding acetolactate synthase 3, was found to be highly expressed. Acetolactate synthase catalyzes the formation of acetolactate from pyruvate, the first step in the synthesis of the branched-chain amino acids (valine, leucine, and isoleucine). KEGG analysis of DEGs in cluster 12 revealed that these genes were associated with the terms "ribosome", "ribosome biogenesis in eukaryotes", "pyruvate metabolism", "fatty acid elongation", "biosynthesis of unsaturated fatty acids", "fatty acid degradation", "fatty acid metabolism", "peroxisome", "glycolysis/gluconeogenesis", and "taurine and hypotaurine metabolism" (Figures 3C, D). *Glyma.05G167700* in cluster 12 is a homolog of *acetolactate synthase small subunit 2*, which encodes the regulatory subunit of acetohydroxy-acid synthase and is involved in the feedback inhibition by branched-chain amino acids. The regulatory subunit is required for full enzymatic activity and contains two repeats of ~180 amino acids, each of which is able to partially activate the catalytic subunit. Leucine inhibits the enzyme reconstituted by the first repeat, though isoleucine and valine do not, and no branched-chain amino acid inhibits enzymes reconstituted by the second repeat (Lee and Duggleby, 2001; Lee and Duggleby, 2002).

In addition, DEGs in cluster 2 were mainly enriched in genes for phenylpropanoid biosynthesis, protein processing in endoplasmic reticulum, glycerolipid metabolism, endocytosis, starch and sucrose metabolism, pentose and glucuronate interconversions or phenylpropanoid biosynthesis, glycerolipid metabolism, linoleic acid metabolism, and valine, leucine and isoleucine degradation. *Glyma.03G205700* in cluster 2 is a homolog of *ANT1* in *Arabidopsis thaliana*, and gene that encodes the protein aromatic and neutral amino acid transporter 1 (ANT1). Aromatic and neutral amino acids are translocated by ANT1, including tryptophan, tyrosine, histidine, phenylalanine, valine, proline, glutamine, leucine, and arginine (Chen et al., 2001). KEGG analysis of DEGs in cluster 4 showed that these genes are involved in pyruvate metabolism, phenylalanine metabolism, phenylpropanoid biosynthesis, pentose and glucuronate interconversions, phenylalanine, tyrosine and tryptophan biosynthesis, tyrosine metabolism, the pentose phosphate pathway, and α-Linolenic acid metabolism (Figure S4). *Glyma.04G236900* in cluster 4 was annotated as glutamate synthase [NADH], and this enzyme is involved in glutamate biosynthesis. These results suggest that the DEGs are related to hormone signal transduction, secondary metabolite

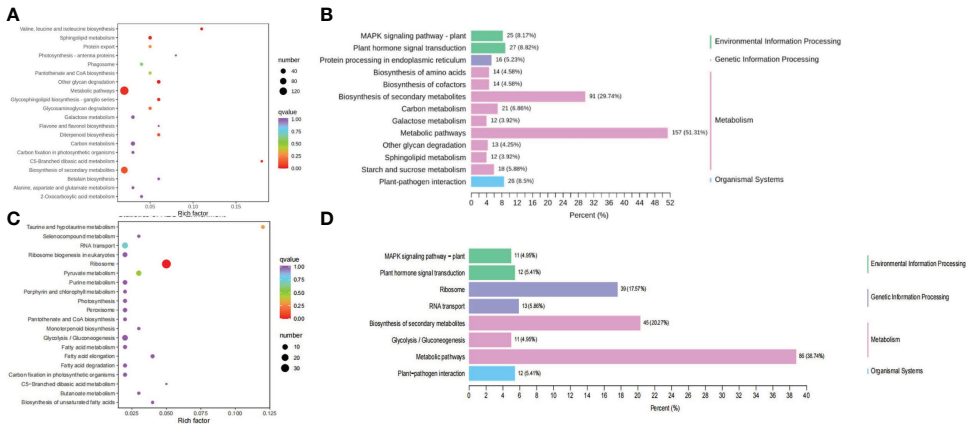


FIGURE 3
KEGG analysis of the DEGs from cluster 3 and 12. **(A)** KEGG classification of the DEGs from cluster 3. **(B)** KEGG enrichment analysis of the DEGs from cluster 3. **(C)** KEGG classification of the DEGs from cluster 12. **(D)** KEGG enrichment analysis of the DEGs from cluster 12.

biosynthesis or phenylpropanoid biosynthesis, amino acid biosynthesis, and glycerolipid metabolism.

DEGs involved in FA biosynthesis

To evaluate the expression of FA biosynthetic genes, we investigated the expression patterns of DEGs involved in lipid biosynthesis and signal transduction pathways (Figure 4). Most genes related to the *de novo* FA biosynthesis pathway were up-regulated in NPS301 during the first three seed developmental periods (7, 14, and 21 DAF), including the subunits of acetyl-CoA carboxylases and the FA synthase complex. Notably, several genes were down-regulated in NPS301, including one gene for fatty acyl thioesterase B (FATB) (*Glyma.04G151600*), two acyl-CoA binding protein genes (*Glyma.11G014900* and *Glyma.13G152900*), and one oleosin gene (*Glyma.19G004800*). Also, the expression of one ketoacyl-CoA synthase gene (*Glyma.06G214800*) was slightly down-regulated in NPS301. Moreover, a gene encoding ketoacyl-ACP synthase I (*Glyma.08G024700*) was up-regulated >5-fold in NPS301-2 versus NPS233-2, and an acyl carrier protein gene (*Glyma.20G230100*) was highly expressed in the latter three developmental periods (14, 21, and 28 DAF) in NPS301.

DEGs involved in the Krebs cycle and amino acid biosynthesis

To identify putative genes involved in the regulation of protein and oil contents in developing soybean seeds, we analyzed the expression of Krebs cycle genes and the metabolic network within which it is embedded. As a result, most of the genes that encode Krebs cycle enzymes had higher expression levels in NPS301 than in NPS233 (Figure 5). With regard to the reactions that consume or produce TCA cycle intermediates, we investigated the expression of genes that encode enzymes involved in amino acid biosynthesis and degradation, secondary metabolite biosynthesis, and fatty acid elongation. Several genes that putatively function in amino acid biosynthesis pathways were up-regulated in NPS233 (Figure 5); and example is *Glyma.16G041200*, which encodes glutamate dehydrogenase 1. *Glyma.02G014800* was annotated as a gene encoding bifunctional aspartate aminotransferase and glutamate/aspartate-prephenate aminotransferase (an *AtAAT* homolog). The *AtAAT* gene is required for the transamination of prephenate to arogenate, and is involved in the aromatic amino acids biosynthesis pathway (Torre et al., 2006). This gene was up-regulated in NPS233 compared to NPS301 in the first developmental period, but down-regulated in the fourth period. Another gene, *Glyma.14G111800*, encodes a homolog of the *aspartate aminotransferase P2* gene in *Lupinus angustifolius*. In *Arabidopsis thaliana*, this gene is related to nitrogen metabolism as well as energy and carbon metabolism, and is important for the metabolizing organic acids related to the Krebs cycle and amino acids (Schultz and Coruzzi, 1995). We also identified several genes that were down-regulated in NPS233 versus NPS301, including an argininosuccinate lyase gene (*Glyma.06G096700*), two putative branched-chain-amino-acid aminotransferase 7 genes (*Glyma.07G186100* and *Glyma.08G063200*), a glutamate dehydrogenase 2 gene (*Glyma.01G204600*), and a branched-chain-amino-acid aminotransferase-like protein 2 gene (*Glyma.19G237000*).

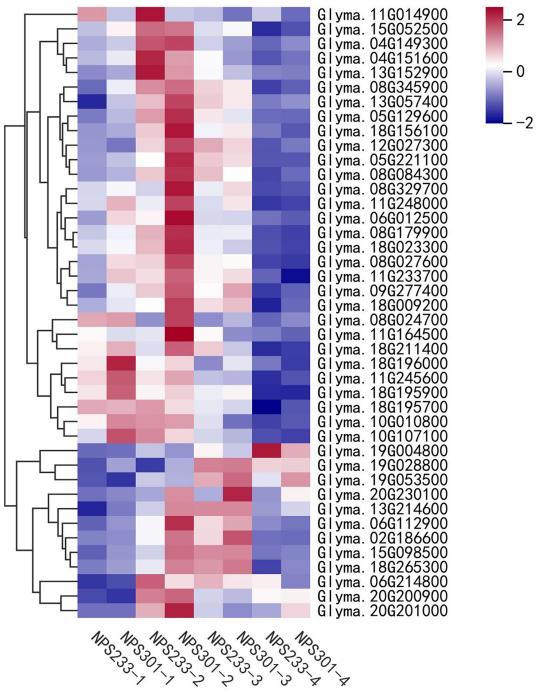


FIGURE 4
Expression patterns of DEGs involved in lipid biosynthesis and signal transduction pathways in soybean seed development. Heatmap showing the expression patterns of the candidate genes involved in the regulation of oil content accumulation in soybean seed. *Glyma.11G014900*, acyl CoA binding protein; *Glyma.15G052500*, hydroxyacyl-ACP dehydrase; *Glyma.04G149300*, ketoacyl-CoA synthase; *Glyma.04G151600*, fatty acyl thioesterase B; *Glyma.13G152900*, acyl CoA binding protein; *Glyma.08G345900*, enoyl-ACP reductase; *Glyma.13G057400*, heteromeric acetyl CoA carboxylase, biotin carboxyl carrier protein; *Glyma.05G129600*, ketoacyl-ACP synthase I; *Glyma.18G156100*, enoyl-ACP reductase; *Glyma.12G027300*, enoyl-ACP reductase; *Glyma.05G221100*, heteromeric acetyl CoA carboxylase, biotin carboxylase subunit; Glyma.08G084300, ketoacyl-ACP synthase I; *Glyma.08G329700*, acyl-CoA synthase; *Glyma.11G248000*, ketoacyl-ACP reductase; *Glyma.06G012500*, ketoacyl-CoA synthase; *Glyma.08G179900*, hydroxyacyl-ACP dehydrase; *Glyma.18G023300*, biotin/lipoyl attachment domain-containing protein; *Glyma.08G027600*, heteromeric acetyl CoA carboxylase, biotin carboxylase subunit; *Glyma.11G233700*, biotin/lipoyl attachment domain-containing protein; *Glyma.09G277400*, ketoacyl-ACP synthase III; *Glyma.18G009200*, ketoacyl-ACP reductase; *Glyma.08G024700*, ketoacyl-ACP synthase I; *Glyma.11G164500*, malonyl CoA-ACP malonyltransferase; *Glyma.18G211400*, ketoacyl-ACP synthase III; *Glyma.18G196000*, heteromeric acetyl CoA carboxylase, carboxyltransferase alpha subunit; *Glyma.11G245600*, ketoacyl-CoA reductase; *Glyma.18G195900*, heteromeric acetyl CoA carboxylase, carboxyltransferase alpha subunit; *Glyma.18G195700*, heteromeric acetyl CoA carboxylase, carboxyltransferase alpha subunit; *Glyma.10G010800*, ER long-chain acyl- CoA synthetase; *Glyma.10G107100*, glycerol 3 phosphate dehydrogenase; *Glyma.19G004800*, Oleosin; *Glyma.19G028800*, heteromeric acetyl CoA carboxylase, biotin carboxyl carrier protein; *Glyma.19G053500*, glycerol 3 phosphate dehydrogenase; *Glyma.20G230100*, acyl carrier protein; *Glyma.13G214600*, acyl carrier protein; *Glyma.06G112900*, plastidic long-chain acyl- CoA synthetase; *Glyma.02G186600*, glycerol 3 phosphate dehydrogenase; *Glyma.15G098500*, acyl carrier protein; Glyma.18G265300, heteromeric acetyl CoA carboxylase, biotin carboxyl carrier protein; *Glyma.06G214800*, ketoacyl-CoA synthase; *Glyma.20G200900*, Caleosins; *Glyma.20G201000*, Caleosins. The gene per row is Z-score standardized.

Discussion

Seed development can be classified into three main stages: the first includes embryo growth, cell division, and morphogenesis; the second includes seed maturation and the accumulation of reserves; and the third includes the desiccation of seeds and subsequent dormancy (Weber et al., 2004). In our study, we performed transcriptomics and metabolomics analyses of soybean seeds in four seed developmental periods. These four developmental periods belong to the first two fundamental stages of seed development.

Metabolites and gene regulatory networks for soybean seed development have been studied in previous reports (Collakova et al., 2013; Peng et al., 2021). More recently, comparative metabolome and transcriptome analyses were performed in the developing seeds of grain and vegetable soybeans at R6 stage, 299 DAMs and and 20,546 DEGs were identified between the two varieties (Chen et al., 2022). Functional enrichment analysis revealed that metabolic pathways, including alanine, aspartate and glutamate metabolism, fatty acid degradation, starch and sucrose metabolism, and flavonoid biosynthesis, were up-regulated in vegetable soybean (Chen et al., 2022), which could partly explain the high-quality of soybean.

The purpose of our study is to investigate the mechanisms that regulate soybean seed oil and protein contents accumulation in developing seeds, our results of enrichment analysis were consistent with the previous report. The DEGs were significantly

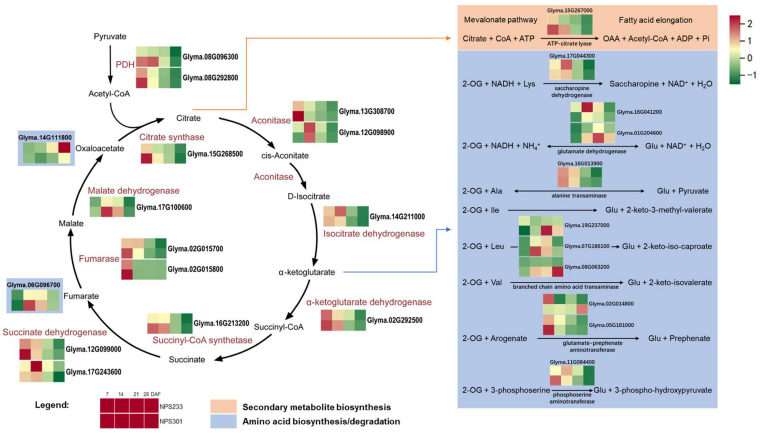


FIGURE 5
DEGs involved in Krebs cycle and the amino acids metabolic network it embedded in. Reactions are shown that consume or produce TCA cycle intermediates. For simplicity, co-enzymes are omitted from the TCA cycle. For each gene, squares denote expression patterns in each variety (see legend). *Glyma.14G111800*, aspartate aminotransferase, Asp + 2-Oxoglutarate <-> Oxaloacetate + Glu, *Glyma.06G096700*, argininosuccinate lyase, Argininosuccinate -> Arg + Fumarate. The gene per row is Z-score standardized.

enriched in pathways related to amino acid and fatty acid metabolism, partially explaining the corresponding differentially accumulated metabolites detected between the two varieties.We performed two separate analyses of differentially expressed genes (DEGs); the first was between different growth periods of the same variety, and the second was between the two varieties at the same growth period. The pairwise comparisons of samples from the four different growth periods identified 12,712 DEGs, and the comparisons of groups from different varieties at the same developmental periods identified 655 DEGs. Among them, there are 461 DEGs common to both comparisons, and 194 genes were exclusively differentially expressed in the comparisons between the two soybean varieties (Figure 6A). We mapped the selected set of genes to the lipid biosynthesis and amino acids biosynthesis pathways to determine their expression patterns. A total of 30 genes were selected, and this candidate DEG set was then used in further analyses (Figure 6B).

Among these 30 genes, two (*Glyma.04G050300* and *Glyma.04G104700*) are involved in leaf senescence delay; *Glyma.04G050300* is annotated as a zinc finger CCCH domain-containing protein 2 (an *OsC3H2* homolog), and *Glyma.04G104700* is annotated as encoding an acyltransferase-like protein (homologous to *At1g54570*). *OsC3H2* may repress the role of jasmonic acid (JA) signaling in promoting leaf senescence and the regulation of panicle development and the pollination/fertilization processes (Kong et al., 2006). Acyltransferase contributes to the synthesis of fatty acid phytyl ester in chloroplasts, which is essential for maintaining the integrity of the photosynthetic membrane during abiotic stress and senescence (Lippold et al., 2012). Expression of both genes was up-regulated in the second seed developmental period in NPS233 versus NPS301, implying that delayed leaf senescence allows more protein to accumulate in the seeds. Previous studies have shown that addition to the *de novo* synthesis of amino acids in seed tissues as they develop an important source of additional free amino acids are those synthesized in vegetative tissues and then transported to seed tissues. The large-scale migration of free amino acids to seeds is related to leaf senescence, where leaves degrade their proteins, producing available free amino acids that can move to develop seeds (Fernie and Hoefgen, 2013; Cohen et al., 2017; Watanabe et al., 2017).

Another gene, *Glyma.17G149300*, is a homolog of *CKB2*. *CKB2* encodes a casein kinase II subunit beta-2, which is involved in the regulation of the basal catalytic activity of the alpha subunit. The tetrameric holoenzyme CK2 has two alpha and two beta subunits and is responsible for phosphorylating the transcription factor PIF1 once it's exposed to light. This induces proteasome-dependent PIF1 degradation and promotes photomorphogenesis (Bu et al., 2011). *Glyma.17G149300* expression was up-regulated in NPS233 versus NPS301 in the first developmental period, indicating that photomorphogenesis in NPS233 occurs earlier than in NPS301.

In addition, the expression of two seed desiccation-related genes, *Glyma.07G236800* and *Glyma.07G236900*, was up-regulated in the third period in NPS301 versus NPS233. These two genes were annotated as encoding desiccation-related protein pcC13-62, and quantitative analysis demonstrated that there is a lower level of pcC13-62 transcript accumulation in species prone to desiccation than in those that tolerate desiccation (Giarola et al., 2018). Synthesizing seed-storage proteins still happens in the third seed developmental stage (maturation and desiccation), though the oil content typically decreased in this stage (Baud et al., 2002). These results indicate that high-protein soybeans may not tolerate desiccation, and thus accumulate more protein.

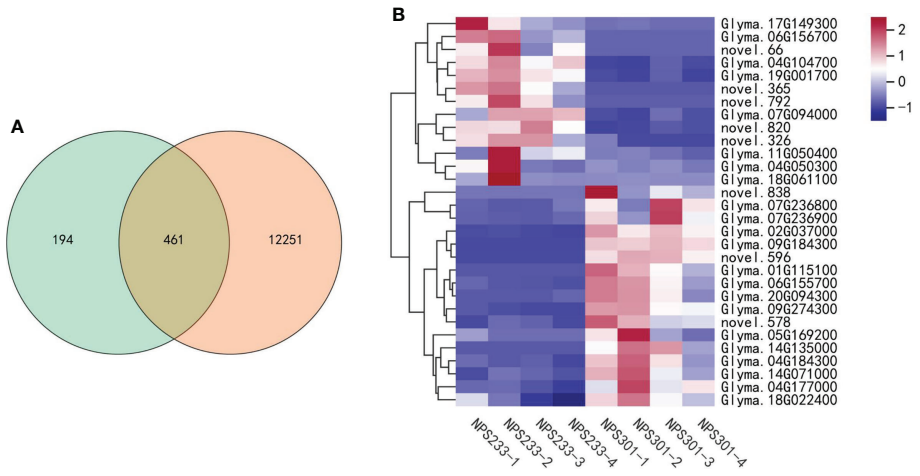


FIGURE 6
Candidate DEG set belonging to the lipid and amino acids biosynthesis pathways. **(A)** Venn diagram depicting the specific DEGs between the four compared groups. **(B)** Heatmap showing the expression patterns of the candidate genes involved in lipid and amino acids biosynthesis pathways. The gene per row is Z-score standardized.

In our study, we observed that the gene encoding the desiccation-tolerance protein pcC13-62 was up-regulated in NPS301; at the same time, the gene *CKB2* that promotes photomorphogenesis was highly expressed in NPS233, and two genes involved in the delay of leaf senescence were also up-regulated in NPS233. Moreover, the two soybean varieties that accumulated different protein and oil contents had different maturity dates, with NPS233 maturing earlier. Taken collectively, we propose that the soybean variety NPS233 is more sensitive to desiccation, photomorphogenesis in NPS233 occurs earlier, and leaf senescence is delayed, which could explain why this variety has a higher seed protein content.

There was also a proteasome-related gene identified in the candidate gene set. the 26S proteasome is a protein complex that is responsible for selective, efficient, and processive hydrolysis of intracellular proteins. *Glyma.14G071000*, annotated as encoding the 26S proteasome regulatory subunit 10B homolog A, was up-regulated in NPS301 compared to NPS233. The 26S proteasome is related to the ATP-dependent degradation of ubiquitinated proteins. Few compounds are used to transport and store most nitrogen in plants. The most common compound related to transport in legumes is asparagine (Miflin et al., 1977; Watanabe et al., 1977). An asparagine synthetase gene, *Glyma.18G061100*, was also present in the candidate gene set, and the gene was up-regulated in NPS233 compared to NPS301 in the seed developmental second period, indicating that there is more nitrogen transport in the high protein/low oil soybean variety NPS233.

It is worth noting that in our study, a gene annotated as encoding glutelin type-A 2 (GLUA2), was up-regulated in NPS301 compared to NPS233. *OsGluA2* is involved in the regulation of grain protein content in rice (Yang et al., 2019); it functions as a positive regulator of rice grain protein content and has a pleiotropic effect on rice grain quality. The grain total protein content, as well as the glutelin, albumin, and prolamin contents, is significantly higher in the high grain protein content accession, but globulin is the exception. However, in soybean, globulin is the predominant seed storage protein. The results of our study show that GLUA2 may function differently in soybean compared to rice. The candidate genes used in this study can be manipulated through gene editing or molecular marker-aided selection to improve the quality of soybeans.

One key outcome of our study is a set of potential key candidate genes. Indeed, within the candidate genes are many fatty acid and amino acid metabolism-related genes, including ten acetyl-CoA carboxylase encoding genes, which initiated the *de novo* FA biosynthesis pathway. Metabolites with different pattern of accumulation between the two varieties with respect to amino acid metabolism were identified. For example, α-ketoglutaric acid, was up-regulated in 'NPS233' versus 'NPS301' in developmental stage 2, whose carboxylate, 2-oxoglutarate provides carbon skeleton for Glu, Gln, proline (Pro) and Arg biosynthesis, and is an intermediate in the Krebs cycle. Some metabolites were not detected, possibly due to the method used in this study, their roles in oil and protein contents accumulation of soybean seed remain to be studied.

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

HC designed and supervised this research. WX performed the experiments and obtained transcriptome and metabolome data. QW analyzed the data. QW and WX prepared the draft of the manuscript. WX, WZ, HZ, XL, XYC, QS, YZ and XC revised and improved the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the National Key Research and Development Program of China (2018YFE0112200), the Key R&D project of Jiangsu Province (BE2019376), the Jiangsu Agriculture Science and Technology Innovation Fund (JASTIF) CX (20)2007, and The Open Competition Project of Seed Industry Revitalization of Jiangsu Province (JBGS [2021]060).

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.1012394/full#supplementary-material

SUPPLEMENTARY FIGURE 1
KEGG enrichment analysis of the differentially accumulated metabolites through the comparative analysis.

SUPPLEMENTARY FIGURE 2
KEGG enrichment analysis of the differentially expressed genes in NPS233 and NPS301.

SUPPLEMENTARY FIGURE 3
Dynamics of gene expression during soybean seed development.

SUPPLEMENTARY FIGURE 4
KEGG classification and enrichment analysis of the differentially expressed genes from the K-means cluster 2 and 4.

References

Bates, P. D. (2016). Understanding the control of acyl flux through the lipid metabolic network of plant oil biosynthesis. *Biochim. Biophys. Acta (BBA) - Mol. Cell Biol. Lipids* 1861, 1214–1225. doi: 10.1016/j.bbalip.2016.03.021

Baud, S., Boutin, J. P., Miquel, M., Lepiniec, L., and Rochat, C. (2002). An integrated overview of seed development in *Arabidopsis thaliana* ecotype WS. *Plant Physiol. Biochem.* 40 (2), 151–160. doi: 10.1016/S0981-9428(01)01350-X

Bhati, K. K., Riyazuddin, R., Pathak, A. K., and Singh, A.(2021). The survey of genetic engineering approaches for oil/fatty acid content improvement in oilseed crops. *Genome Eng. Crop Improv*. 22–42. doi: 10.1002/9781119672425.ch11

Bremner, J. M. (1960). Determination of nitrogen in soil by the kjeldahl method. *J. Agric. Sci.* 55 (1), 11–33. doi: 10.1017/S0021859600021572

Bu, Q., Zhu, L., Dennis, M. D., Yu, L., Lu, S. X., Person, M. D., et al. (2011). Phosphorylation by CK2 enhances the rapid light-induced degradation of phytochrome interacting factor 1 in *Arabidopsis*. *J. Biol. Chem.* 286 (14), 12066. doi: 10.1074/jbc.M110.186882

Chaudhary, J., Patil, G., Sonah, H., Deshmukh, R., Vuong, T., Valliyodan, B., et al. (2015). Expanding omics resources for improvement of soybean seed composition traits. *Front. Plant Sci.* 6 (31)1021. doi: 10.3389/fpls.2015.01021

Chen, L., Ortiz-Lopez, A., Jung, A., and Bush, D. R.. (2001). ANT1, an aromatic and neutral amino acid transporter in *Arabidopsis*. *Plant Physiol.* 125 (4), 1813–1820. doi: 10.1104/pp.125.4.1813

Chen, Z., Zhong, W., Zhou, Y., Ji, P., Wan, Y., Shi, S., et al. (2022). Integrative analysis of metabolome and transcriptome reveals the improvements of seed quality in vegetable soybean (Glycine max (L.) merr.). *Phytochemistry* 200, 113216. doi: 10.1016/j.phytochem.2022.113216

Chung, J., Babka, H. L., Graef, G. L., Staswick, P. E., Lee, D. J., Cregan, P. B., et al. (2003). The seed protein, oil, and yield QTL on soybean linkage group I. *Crop Sci.* 43 (3), 1053–1067. doi: 10.2135/cropsci2003.1053

Clemente, T. E., and Cahoon, (2009). Soybean oil: Genetic approaches for modification of functionality and total content. *E. B. Plant Physiol.* 151 (3), 1030–1040. doi: 10.1104/pp.109.146282

Cohen, H., Hacham, Y., Panizel, I., Rogachev, I., Aharoni, A., Amir, R., et al. (2017). Repression of CYSTATHIONINE gamma-SYNTHASE in seeds recruits the s-methylmethionine cycle. *Plant Physiol.* 174, 1322–1333. doi: 10.1104/pp.17.00579

Collakova, E., Aghamirzaie, D., Fang, Y., Klumas, C., Tabataba, F., Kakumani, A., et al. (2013). Metabolic and transcriptional reprogramming in developing soybean (Glycine max) embryos. *Metabolites* 3 (2), 347–372. doi: 10.3390/metabo3020347

Fehr, W. R., Collins, F. I., and Weber, C. R. (1968). Evaluation of methods for protein and oil determination in soybean seed. *Crop Sci.* 8 (1). doi: 10.2135/cropsci1968.0011183X000800010015x

Giarola, V., Jung, N. U., Singh, A., Satpathy, P., and Bartels, D. (2018). Analysis of pcC13-62 promoters predicts a link between cis-element variations and desiccation tolerance in linderniaceae. *J. Exp. Bot.* 69 (15), 3773–3784. doi: 10.1093/jxb/ery173

Götz, S., García-Gómez, J. M., Terol, J., Williams, T. D., Nagaraj, S. H., Nueda, M. J., et al. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res.* 36, 3420–3435. doi: 10.1093/nar/gkn176

Kirma, M., Araújo, W. L., Fernie, A. R., and Galili, G. (2012). The multifaceted role of aspartate-family amino acids in plant metabolism. *J. Exp. Bot.* 63 (14), 4995–5001. doi: 10.1093/jxb/ers119

Kong, Z., Li, M., Yang, W., Xu, W., and Xue, Y. (2006). A novel nuclear-localized CCCH-type zinc finger protein, OsDOS, is involved in delaying leaf senescence in rice. *Plant Physiol.* 141, 1376–1388. doi: 10.1104/pp.106.082941

Lee, Y. T., and Duggleby, R. G. (2001). Identification of the regulatory subunit of *Arabidopsis thaliana* acetohydroxyacid synthase and reconstitution with its catalytic subunit. *Biochemistry* 40 (23), 6836–6844. doi: 10.1021/bi002775q

Lee, Y. T., and Duggleby, R. G. (2002). Regulatory interactions in *Arabidopsis thaliana* acetohydroxyacid synthase. *FEBS Lett.* 512 (1-3), 180–184. doi: 10.1016/S0014-5793(02)02253-6

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 (6), 1639–1659. doi: 10.1007/s00122-019-03304-5

Less, H., and Galili, G. (2008). Principal transcriptional programs regulating plant amino acid metabolism in response to abiotic stresses. *Plant Physiol.* 147 (1), 316–330. doi: 10.1104/pp.108.115733

Less, H., and Galili, G. (2009). Coordinations between gene modules control the operation of plant amino acid metabolic networks. *BMC Syst. Biol.* 3 (1), 14. doi: 10.1186/1752-0509-3-14

Li, L., Hur, M., Lee, J. Y., Zhou, W., Song, Z., Ransom, N., et al. (2015). A systems biology approach toward understanding seed composition in soybean. *BMC Genomics* 16 (3), S9. doi: 10.1186/1471-2164-16-S3-S9

Li, D., Zhao, X., Han, Y., Li, W., and Xie, F.. (2018). Genome-wide association mapping for seed protein and oil contents using a large panel of soybean accessions. *Genomics* 111 (1), 90–95. doi: 10.1016/j.ygeno.2018.01.004

Lippold, F., Drop, K. V., Abraham, M., Hölzl, G., Wewer, V., Yilmaz, J. L., et al. (2012). Fatty acid phytyl ester synthesis in chloroplasts of *Arabidopsis*. *Plant Cell* 24 (5), 2001–2014. doi: 10.1105/tpc.112.095588

Lu, D., Huang, Z., Liu, D., Zhu, P., Lv, S., Li, N., et al. (2018). Transcriptome analysis of chrysanthemum in responses to white rust. *Scientia. Hortic.* 233, 421–430. doi: 10.1016/j.scienta.2018.01.016

Mao, X., Cai, T., Olyarchuk, J. G., and Wei, L. (2005). Automated genome annotation and pathway identification usingthe KEGG orthology (KO) as a controlled vocabulary. *Bioinformatics* 21, 3787–3793. doi: 10.1093/bioinformatics/bti430

Miflin, B. J., and Lea, P. J. (1977). Amino acid metabolism. *Annu. Rev. Plant Physiol.* 28 (1), 299–329. doi: 10.1146/annurev.pp.28.060177.001503

Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., and Wold, B.. (2008). Mapping and quantifying mammalian transcriptomes by RNA-seq. *Nat. Methods* 5, 621–628. doi: 10.1038/nmeth.1226

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

Peng, L., Qian, L., Wang, M., Liu, W., Song, X., Cheng, H., et al. (2021). Comparative transcriptome analysis during seeds development between two soybean cultivars. *PeerJ* 9, e10772. doi: 10.7717/peerj.10772

Pratelli, R., and Pilot, G. (2014). Regulation of amino acid metabolic enzymes and transporters in plants. *J. Exp. Bot.* 65 (19), 5535–5556. doi: 10.1093/jxb/eru320

Salie, M. J., and Thelen, J. J. (2016). Regulation and structure of the heteromeric acetyl-CoA carboxylase. *Biochimica et biophysica acta (BBA). Mol. Cell Biol. Lipids* 1861 (9 Pt B), 1207–1213. doi: 10.1016/j.bbalip.2016.04.004

Schultz, C. J., and Coruzzi, G. M. (1995). The aspartate aminotransferase gene family of *Arabidopsis* encodes isoenzymes localized to three distinct subcellular compartments. *Plant J.* 7, 61–75. doi: 10.1046/j.1365-313X.1995.07010061.x

Torre, F., Santis, L. D., Suárez, M. F., Crespillo, R., and Cánovas, F. M. (2006). Identification and functional analysis of a prokaryotic-type aspartate aminotransferase: implications for plant amino acid metabolism. *Plant J. Cell Mol. Biol.* 46 (3), 414–425. doi: 10.1111/j.1365-313X.2006.02713.x

Watanabe, M., Balazadeh, S., Tohge, T., Erban, A., Giavalisco, P., Kopka, J., et al (2013). Comprehensive dissection of spatiotemporal metabolic shifts in primary, secondary, and lipid metabolism during developmental senescence in *Arabidopsis*. *Plant Physiol.* 162 (3), 1290–1310. doi: 10.1104/pp.113.217380

Weber, H., Borisjuk, L., and Wobus, U. (2004). Molecular physiology of legume seed development. *Annu. Rev. Plant Biol.* 56 (1), 253–279. doi: 10.1146/annurev.arplant.56.032604.144201

Xu, C., and Shanklin, J. (2016). Triacylglycerol metabolism, function, and accumulation in plant vegetative tissues. *Annu. Rev. Plant Biol.* 67 (1), 179–206. doi: 10.1146/annurev-arplant-043015-111641

Yang, Y., Guo, M., Sun, S., Zou, Y., Yin, S., Liu, Y., et al. (2019). Natural variation of Os*GluA2* is involved in grain protein content regulation in rice. *Nat. Commun.* 10 (1), 1949. doi: 10.1038/s41467-019-09919-y

Young, G., Mebrahtu, T., and Johnson, J. (2000). Acceptability of green soybeans as a vegetable entity. *Plant Foods Hum. Nutr.* 55 (4), 323–333. doi: 10.1023/A:1008164925103

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 (3), 460–472. doi: 10.1016/j.molp.2017.12.016

Zhang, S. S., Hao, S., Zhang, S., Zhang, D., Wang, H., Du, H., et al. (2021). Genome–wide association mapping for protein, oil and water–soluble protein contents in soybean. *Mol. Genet. Genomics* 296 (1), 91–102. doi: 10.1007/s00438-020-01704-73

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 24 October 2022
DOI 10.3389/fpls.2022.1024543

OPEN ACCESS

EDITED BY
Jianghua Chen,
Key Laboratory of Tropical Plant Resource and Sustainable Use (CAS), China

REVIEWED BY
Eleonora Cominelli,
National Research Council (CNR), Italy
Usman Aziz,
Northwestern Polytechnical University, China

*CORRESPONDENCE
Dinakaran Elango
delango@iastate.edu

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Functional and Applied Plant Genomics, a section of the journal Frontiers in Plant Science

RECEIVED 02 September 2022
ACCEPTED 10 October 2022
PUBLISHED 24 October 2022

CITATION
Elango D, Wang W, Thudi M, Sebastiar S, Ramadoss BR and Varshney RK (2022) Genome-wide association mapping of seed oligosaccharides in chickpea. *Front. Plant Sci.* 13:1024543. doi: 10.3389/fpls.2022.1024543

Genome-wide association mapping of seed oligosaccharides in chickpea

Dinakaran Elango [1,2*†], Wanyan Wang [3†], Mahender Thudi [4,5,6], Sheelamary Sebastiar [7], Bharathi Raja Ramadoss [8] and Rajeev K. Varshney [6,9]

[1]Department of Agronomy, Iowa State University, Ames, IA, United States, [2]Department of Plant Science, Penn State University, University Park, PA, United States, [3]Ecosystem Science and Management, Penn State University, University Park, PA, United States, [4]Department of Agricultural Biotechnology and Molecular Biology, Dr. Rajendra Prasad Central Agricultural University, Samastipur, India, [5]Centre for Crop Health, University of Southern Queensland (USQ), Toowoomba, QLD, Australia, [6]Genetics Gains Research Program, International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Hyderabad, India, [7]Division of Crop Improvement, Indian Council of Agricultural Research (ICAR)-Sugarcane Breeding Institute, Coimbatore, India, [8]Saskatoon Research and Development Centre, Agriculture and Agri-Food Canada, Saskatoon, SK, Canada, [9]State Agricultural Biotechnology Centre, Crop Research Innovation Centre, Food Futures Institute, Murdoch University, Murdoch, WA, Australia

Chickpea (*Cicer arietinum* L.) is one of the major pulse crops, rich in protein, and widely consumed all over the world. Most legumes, including chickpeas, possess noticeable amounts of raffinose family oligosaccharides (RFOs) in their seeds. RFOs are seed oligosaccharides abundant in nature, which are non-digestible by humans and animals and cause flatulence and severe abdominal discomforts. So, this study aims to identify genetic factors associated with seed oligosaccharides in chickpea using the mini-core panel. We have quantified the RFOs (raffinose and stachyose), ciceritol, and sucrose contents in chickpea using high-performance liquid chromatography. A wide range of variations for the seed oligosaccharides was observed between the accessions: 0.16 to 15.13 mg g^{-1} raffinose, 2.77 to 59.43 mg g^{-1} stachyose, 4.36 to 90.65 mg g^{-1} ciceritol, and 3.57 to 54.12 mg g^{-1} for sucrose. Kabuli types showed desirable sugar profiles with high sucrose, whereas desi types had high concentrations RFOs. In total, 48 single nucleotide polymorphisms (SNPs) were identified for all the targeted sugar types, and nine genes (*Ca_06204, Ca_04353,* and *Ca_20828*: *Phosphatidylinositol N-acetylglucosaminyltransferase*; *Ca_17399* and *Ca_22050*: *Remorin proteins*; *Ca_11152*: *Protein-serine/threonine phosphatase; Ca_10185, Ca_14209*, and *Ca_27229*: *UDP-glucose dehydrogenase*) were identified as potential candidate genes for sugar metabolism and transport in chickpea. The accessions with low RFOs and high sucrose contents may be utilized in breeding specialty chickpeas. The identified candidate genes could be exploited in marker-assisted breeding, genomic selection, and genetic engineering to improve the sugar profiles in legumes and other crop species.

KEYWORDS

anti-nutritional factors (ANF), flatus potential, marker trait associations, prebiotics, raffinose family oligosaccharides (RFOs), specialty chickpeas

Introduction

Chickpea (*Cicer arietinum* L.) is one of the founder crops domesticated between 9,000–11,000 years ago and is an important ancient legume grown and consumed all over the globe (Lev-Yadun et al., 2000). As a legume crop, it is often grown as rotational crops with cereals to enhance yield because of their ability to fixing atmospheric nitrogen (Graham and Vance, 2003). Chickpea is rich in carbohydrates (60-65%), protein (20-22%), fat (6%), and rich in dietary fiber, as well as minerals (phosphorus, calcium, magnesium, iron, and zinc) and vitamins (β-carotene, thiamin, riboflavin, and niacin) (Jukanti et al., 2012). The major pulse grain constituents are carbohydrates, based on their polymeric structure that can be classified as monosaccharides (ribose, fructose, and glucose), disaccharides (sucrose, maltose, melibiose), oligosaccharides (raffinose, stachyose, verbascose, ajugose, and ciceritol) and polysaccharides (Chibbar et al., 2010). Among the oligosaccharides, *α-galacto-oligosaccharides* (*α-GOS*) are known as raffinose family oligosaccharides (RFOs) (Sosulski et al., 1982). The major RFOs found in chickpea include raffinose, stachyose, and verbascose. However, ciceritol does not belong to the RFOs since its structure is different from *α-GOS* and can rapidly undergo a hydrolysis process, so unlike raffinose and stachyose, ciceritol does not cause flatulence in humans and animals (Quemener and Brillouet, 1983).

RFOs are the single most deterrent factor for the rapid adoption of legumes in mainstream food usage in humans and animals (Delumen, 1992; Elango et al., 2022a). Humans and animals lack the enzyme *α-galactosidase* to degrade *α-galactosides* (RFOs), which results in the accumulation of undigested RFOs in the large intestine of the digestive system, which ultimately causes flatulence and abdominal discomforts due to the production of flatulent gases by colonic bacteria through fermenting the un-digested RFOs present in the guts (Calloway and Murphy, 1968; Sosulski et al., 1982; Singh, 1985; Han and Baik, 2006). Though, RFOs have been reported to have a beneficial effect on gut microflora (Van den Ende, 2013; Su et al., 2019), and play a role in seed germinability and biotic and abiotic stress tolerance in crop plants (Gulewicz et al., 2002; Taji et al., 2002; Nishizawa-Yokoi et al., 2008; Dobrenel et al., 2013; Van den Ende, 2013; Gangl and Tenhaken, 2016; Yan et al., 2022). However, we do not know what the right concentration is needed to benefit humans, animals, and plants concerning RFOs.

Food processing can eliminate RFOs at varying degrees and significantly increase dietary fraction availability in food (Jood et al., 1985; Egbe and Akinyele, 1990; Aguilera et al., 2009). However, these techniques often come with trade-offs; most techniques are time-consuming, lead to loss of nutrients, and sometimes have consumer acceptability issues. Therefore, identifying sources of variation for developing desirable sugar-type cultivars through crop breeding is very important. Screening and identification of low RFOs have been carried out in many economically important legume crops such as lentil (Tahir et al., 2011), chickpea (Raja et al., 2015; Gangola et al., 2016), pea (Peterbauer et al., 2003), soybean (Blackman et al., 1992; Dierking and Bilyeu, 2008; Obendorf and Górecki, 2012), mung bean and urd bean (Souframanien et al., 2014), whereas, very limited efforts have been taken toward the identification of genomic regions associated with RFOs in crop plants. In this context, our study aims to identify the genetic factors responsible for seed oligosaccharides in chickpea through genome-wide association mapping using the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) mini-core collection.

Materials and methods

Plant materials

The chickpea mini-core collection consisting of 211 accessions from 24 countries (Asia, Africa, Europe, North, and South American regions) was obtained from the genetic resources division of International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), India (Upadhyaya and Ortiz, 2001). Field experiments were performed in a randomized complete block design (RCBD) with three replications in the 2010 winter season at the Department of Pulses (11.0232° N latitude, 76.9293° E longitude, 426.72 m altitude), Tamil Nadu Agricultural University (TNAU), Coimbatore, India. Each accession was grown in a single row in a 3 m long plot. Seeds from each replicate of individual accessions were harvested at physiological maturity and stored at 4°C until analysis was performed. A standard agronomic package of practices was followed to achieve the best crop establishment.

Sugar extraction and quantification

The seeds of each chickpea accession were grounded, and the flours were used to extract soluble sugars. One gram of flour samples was taken into a screw cap vial and mixed with 10 ml of 50% ethanol, and vortexed briefly. After adding ethanol, samples were shaken horizontally using a water bath shaker maintained at 50° C for 30 min at 100 rpm. The incubated vial was centrifuged at 4000 rpm for 5 min, then 5 ml of supernatant was taken and mixed with 7 ml of acetonitrile (high-performance liquid chromatography (HPLC)-grade) to precipitate the soluble proteins. The mixture (5 ml supernatant + 7 ml acetonitrile) was incubated at room temperature for two hours. After incubation, the mixture was centrifuged at 3670 g for 5 min, and one ml aliquot of the supernatant was collected. The collected supernatant was dried at 50°C and resuspended with 500 µl 65% HPLC-grade acetonitrile and filtered through a 0.2 µm membrane filter and transferred to HPLC vials. Standards of sucrose, raffinose, and stachyose were purchased from Sigma-Aldrich, Bengaluru, India and ciceritol was purchased

from Clearsynth, Hyderabad, India. Three different concentrations, 1.25 mg ml^{-1}, 2.5 mg ml^{-1}, and 5.0 mg ml^{-1}, were prepared and included in each batch of samples to obtain the standard curve. The concentration of different sugars (sucrose, raffinose, stachyose, and ciceritol) was determined using the HPLC (Shimadzu, Kyoto, Japan), which consisted of an LC20AD pump and a RID-10A refraction index detector. Sugar concentrations were determined using the peak area of the sample in comparison with standards.

Marker-trait association analyses

We performed genome-wide association mapping analysis using 673,115 single nucleotide polymorphisms (SNPs), where the SNP calls for 211 genotypes were obtained from Varshney et al. (2019). As reported in Varshney et al. (2019), for calling SNPs, the clean reads were mapped on to the reference genome of chickpea genotype CDC Frontier using SOAP2. We then used SOAPsnp3 to calculate the likelihood of all possible genotypes for each sample. In order to filter out low-quality variants, the loci with sequencing depth higher than 10,000 and lower than 400, mapping times higher than 1.5, or quality scores lower than 20, were filtered out. The loci with estimated allele frequency not equal to 0 or 1 were determined as SNPs. After obtaining the SNPs, we also determined the genotype of each individual at the SNP locus by assigning the most likely genotype from the SOAPsnp3 result of each sample. We have used the Fixed and random model Circulating Probability Unification (FarmCPU) model (Liu et al., 2016) in the Genome Association and Prediction Integrated Tool (GAPIT3) package (Wang and Zhang, 2021) to identify significant marker-trait associations (MTAs). GAPIT estimated the allelic effect for the significant SNPs identified. Sign (+/-) of the allelic effect estimate is relative to the alphabetical order of the nucleotides. MTAs were selected for p-value $<10^{-5}$. Gene annotations were determined from the reference genome of chickpeas released in 2013 (Varshney et al., 2013). Genes within the flanking regions of 50kb upstream and downstream of the significantly called SNPs were collected first, and among them, only the genes already annotated in the chickpea genome with predicted function or found orthologs in other model plants were selected as candidate genes. Distance from the SNP (bp) is calculated as the distance from the SNP location to the start site of the upstream or downstream genes. If the SNP is located within a gene, the distance is calculated as the distance to the start site of the gene.

Results

Phenotypic variation and correlations among seed oligosaccharides

We have observed wide variations for all sugars measured in the ICRISAT chickpea mini-core collection. Among the morphotypes, Kabuli-type chickpeas exhibited higher sucrose and total sugar contents. In contrast, the desi-type chickpeas showed higher RFOs (raffinose and stachyose) and ciceritol contents in seeds (Figure 1). Mini-core collection showed wide seed oligosaccharide variations: 0.16 to 15.13 mg g^{-1}, 2.77 to 59.43 mg g^{-1}, 4.36 to 90.65 mg g^{-1}, 3.57 to 54.12 mg g^{-1} for raffinose, stachyose, ciceritol, and sucrose with an average of 4.61, 28.02, 34.48, and 23.11 mg g^{-1} flour sample, respectively (Table 1). We have observed significant positive correlations among seed oligosaccharides measured: sucrose, ciceritol, and stachyose have strong positive correlations with total sugars; moderate correlations were observed between sucrose and ciceritol, and ciceritol and stachyose (Figure 2). Whereas raffinose had a low level of positive correlations with all other seed oligosaccharides (Figure 2).

Marker trait associations

Genome-wide association mapping identified 48 SNPs associated with the seed oligosaccharide contents in chickpeas (Table 2). The largest number of associated markers (12) were detected on chromosome 4 (Table 2), and there were 16 SNPs found to be significantly associated with raffinose content in chickpea, which is the most associated markers compared with other three sugars: 7 SNPs for ciceritol and stachyose respectively, and 9 SNPs for sucrose, and 8 SNPs for total sugar content in chickpea (Table 2; Figure 3).

Probable candidate genes

We identified 80 probable candidate genes for all the seed oligosaccharides measured in this study (Table 2). Among them, nine genes were recognized with annotated functions highly associated with sugar biosynthesis and transportation. Four genes that are important components in inositol biosynthesis: two associated with stachyose (*Ca_06204: phosphatidylinositol N-acetylglucosaminyltransferase; Ca_04353: Type I inositol 1,4,5-trisphosphate 5-phosphatase*), and two associated with sucrose (*Ca_20828: inositol-polyphosphate phosphatase; Ca_10185*: *UDP-glucose dehydrogenase*) have been identified as probable candidates for seed oligosaccharide biosynthesis in chickpeas (Table 2). There were also three candidate genes identified playing an important role in RFOs transportation: two of the genes (*Ca_17399* and *Ca_22050*) are associated with raffinose content, and the other gene (*Ca_11152*) is linked with total sugar content in chickpea seeds, and all of them encodes *remorin protein*, which regulates the carbohydrate translocation in plants. The other group of important genes identified in our study includes *Ca_14209* (associated with two traits: ciceritol and total sugar contents) and *Ca_27229* (associated with

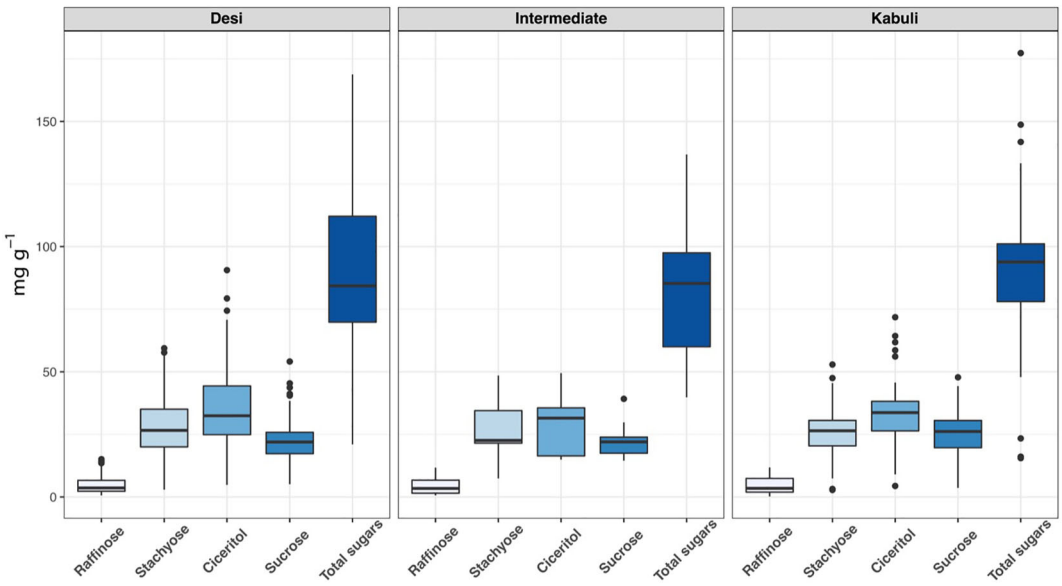


FIGURE 1
Seed oligisaccharides variations among desi, kabuli, and intermediate types in the ICRISAT chickpea mini-core panel.

raffinose content), and both genes encode *UDP-glucose dehydrogenase* (*UGD*) and likely interfere with RFOs biosynthesis as a competitor for the upstream precursor compound (Joët et al., 2009).

A candidate gene *Ca_23689* harboring SNP Ca4_43438450 was associated with raffinose content, and GO annotation indicates gene *Ca_23689* encodes structural constituent of the cell wall (Table 2). A precedent study has discovered that over-expression of *raffinose synthase* (*rfs*) resulted in increased biomass and total cellulose content in the cell wall (Unda et al., 2017). Gene *Ca_03642* was associated with chickpea stachyose content in seed (Table 2). The GO term functional annotation suggests that gene *Ca_03642* involves intracellular protein transport, vesicle-mediated transport, and membrane fusion. Studies have demonstrated that stachyose is the primary photoassimilate and transport sugar in legumes (Peterbauer et al., 2001; Qiu et al., 2015). Another gene, *Ca_09762*, which is predicted as involved in calcium ion transmembrane transport, was also associated with stachyose content (Table 2). Meanwhile, we identified another SNP locus (Ca6_2510863) in the gene *Ca_10383* associated with sucrose content in chickpea seed, and the functional prediction of gene *Ca_10383* is ATP hydrolysis coupled proton transport. A previous study suggested that stachyose and sucrose may also be accumulated in the vacuole by stachyose and Sucrose/H^+ antiporter mechanisms, which is an ATPase energized vacuolar uptake process (Keller, 1992).

Two SNP loci (Ca_46225454 in gene *Ca_21541* and Ca5_1870839 in gene *Ca_26715*) were associated with ciceritol content in chickpea seeds (Table 2). Gene *Ca_21541* is predicted as a *TPX2* (targeting protein for Xklp2) family protein. SNP Ca_46225454 is also adjacent to a sequence fragment ortholog of *AT5G40690*, which encodes for methyltransferase activity. Ciceritol is the end product of the inositol methylation

TABLE 1 Variability of chickpea accessions for seed oligosaccharide contents.

Seed oligosaccharides	Range	Mean ± SD	Desirable sugar profile genotypes mg g^{-1}	Origin	Morphotype
Raffinose	0.16 - 15.13	4.61 ± 3.24	ICC 6263 (0.16)	Russia & CISs	Kabuli
Stachyose	2.77 - 59.43	28.02 ± 11.69	ICC 13816 (2.77)	Russia & CISs	Kabuli
Ciceritol	4.36 - 90.65	34.48 ± 14.38	ICC 12968 (4.36)	India	Kabuli
Sucrose	3.57 - 54.12	23.11 ± 8.32	ICC 12654 (54.12)	Ethiopia	Desi
Total sugars	15.53 - 177.34	90.22 ± 29.93	–	–	–

process, which explains the role of the identified SNP Ca_46225454 in ciceritol biosynthesis in chickpeas. Gene *Ca_26715* is an *endochitinase A-like protein*. Chitinase is known to protect plants against abiotic and biotic stresses (de Las Mercedes Dana et al., 2006; Karlsson and Stenlid, 2008; Xin et al., 2021; Zhang et al., 2022). The association between endochitinase and ciceritol content in chickpea suggests a metabolic link between the ciceritol pathway and the pathway leading to biotic and abiotic resistance. Two SNPs were identified for total sugar content: SNP Ca4_17208915 in gene *Ca_05402* and Ca5_1870839 in gene *Ca_26715*. Gene *Ca_05402* involves ATP hydrolysis coupled proton transport. Gene *Ca_26715* is predicted as an *A-like endochitinase* (Table 2).

Discussion

Breeding specialty chickpeas

Identification of germplasm lines with good nutritional quality parameters is key for developing cultivars for different end users. In this study, we have identified varying sugar profile accessions that could be exploited as a base in breeding specialty chickpeas. Especially the accessions with high sucrose (ICC 12564 and ICC 9137) and low RFOs (ICC 6263 and ICC 13816) are desirable for making delicacies like hummus, besan laddoo, Mysore Pak, besan barfi, Puran Poli, and other confectionaries without refined sugars and or artificial

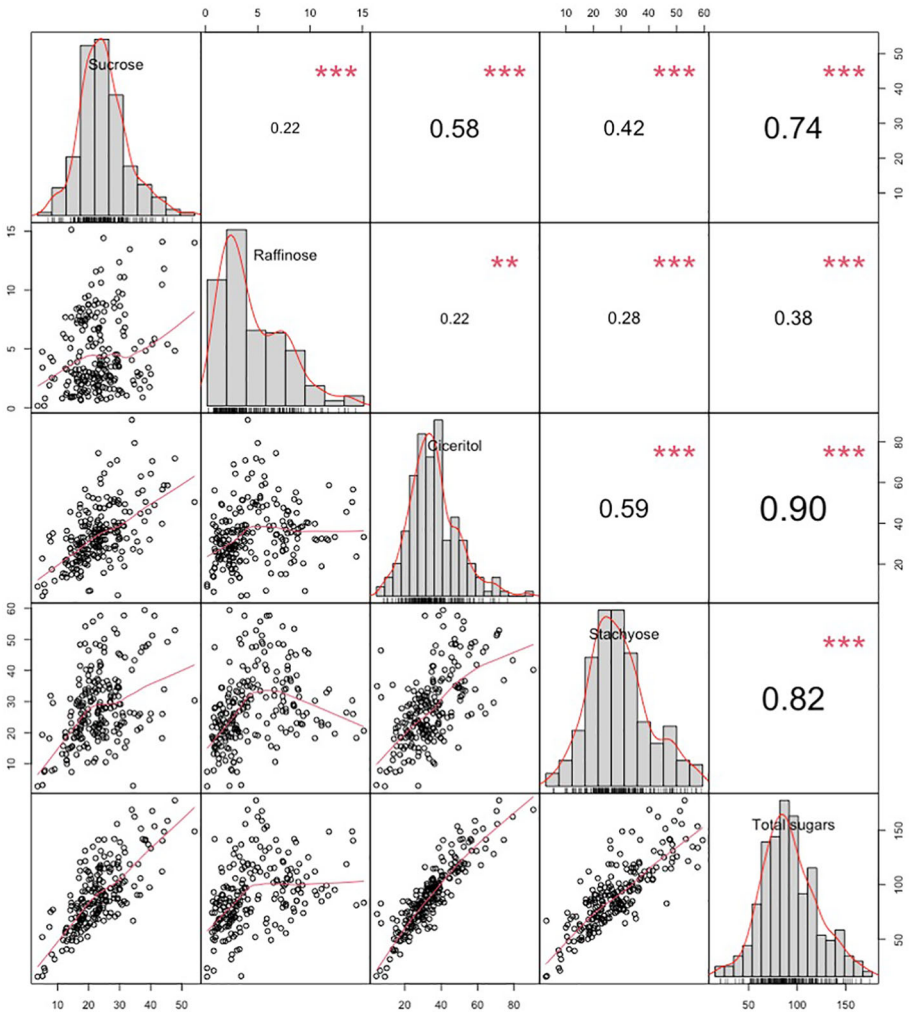


FIGURE 2
Variation and Pearson pairwise correlations of sucrose, raffinose, ciceritol, stachyose, and total sugars in chickpea. Upper diagonal: Pearson correlation coefficients between every two traits. Mid-diagonal: Histograms of sucrose, raffinose, ciceritol, stachyose, and total sugars. Lower diagonal: Bivariate scatter plots of correlations between every two traits with a fitted line. **Significant at the .01 probability level. ***Significant at the .001 probability level.

TABLE 2 List of significant single nucleotide polymorphic associations, the genes tagged by significant single nucleotide polymorphic markers, and candidate genes identified based on proximity to the significant markers and their description for ciceritol, raffinose, stachyose, sucrose, and total sugars in chickpea.

Oligosaccharides	Scaffold position	Allele	*P* value	Minor allele frequency	Allelic effect	Chickpea gene ID	Distance from the SNP (bp)*	SNP position to gene loci	Functional annotation
Ciceritol	***Ca4_30621427***	***A/T***	***2.38E-07***	***0.26***	***-5.54***	***Ca_14209***	***8067***	***Downstream***	***UDP-glucose dehydrogenase***
						Ca_14210	47995	Upstream	TBCC domain-containing protein 1
	Ca5_34185294	C/A	7.22E-07	0.24	7.07	Ca_01829	5411	Downstream	Cryptochrome, DASH family protein
						Ca_01828	1764	Upstream	Predicted membrane protein
	Ca2_15589436	G/C	2.57E-06	0.47	13.91	Ca_18544	47000	Downstream	6-phosphogluconate dehydrogenase
						Ca_18543	19728	Upstream	Receptor-like protein kinase-related
	Ca4_11536839	A/G	5.22E-06	0.43	-6.05	Ca_04384	46920	Downstream	Tubby-like f-box protein 1-related
						Ca_04385	15105	Upstream	Trihelix transcription factor gt-2
	Ca1_46225454	T/C	5.24E-06	0.06	-9.08	Ca_21541	3116	Within the gene	Tpx2 (targeting protein for xklp2) protein family
	Ca5_1870839	G/A	9.02E-06	0.17	-8.41	Ca_26715	68	Within the gene	DNA damage-binding protein 1 (DDB1)
	Ca7_11316700	A/G	9.64E-06	0.28	-6.74	Ca_09363	20294	Downstream	AhpC/TSA antioxidant enzyme (AhpC-TSA_2)
						Ca_09362	7497	Upstream	CAMP-response element binding protein-related
Raffinose	Ca5_42575153	T/A	1.16E-07	0.24	-1.99	Ca_11357	12497	Downstream	F-box only protein 6
						Ca_11356	6613	Upstream	Squamosa promoter-binding-like protein 10-related
	Ca7_19222249	T/C	3.54E-07	0.19	-1.43	Ca_12328	7908	Downstream	Ap2-like ethylene-responsive transcription factor ail6-related
						Ca_12329	6606	Upstream	Tetratricopeptide repeat
	Ca1_36822662	***G/A***	***3.61E-07***	***0.15***	***-2.30***	Ca_21700	29080	Downstream	Ulp1 protease family
						Ca_27229	***17117***	***Upstream***	***Udp-glucose dehydrogenase***
	Ca6_14474273	T/C	9.04E-07	0.12	-1.77	Ca_05217	5850	Downstream	Protein-serine/threonine phosphatase
						Ca_05218	16903	Upstream	Cotton fibre expressed protein
	Ca6_56636684	***T/A***	***1.66E-06***	***0.12***	***-2.37***	***Ca_17399***	***1462***	***Downstream***	***Remorin family protein***
						Ca_17400	1515	Upstream	RNAse P Rpr2/Rpp21/SNM1 subunit domain (Rpr2)
	Ca6_26580794	T/C	2.85E-06	0.25	-1.33	Ca_16690	13777	Downstream	Dof domain, zinc finger (zf-Dof)
						Ca_16691	18135	Upstream	WUSCHEL-related homeobox 2
	Ca4_16178393	G/A	2.89E-06	0.16	-1.59	Ca_05494	6388	Downstream	Zinc finger protein jagged-related
						Ca_05493	4880	Upstream	Protein istr-1, isoform a
	Ca6_18000446	T/G	3.06E-06	0.37	-2.07	Ca_06421	12470	Downstream	Serine/threonine-protein kinase srk2e
						Ca_06422	25071	Upstream	Myb family transcription factor
	Ca7_32073468	C/A	3.55E-06	0.13	-1.75	Ca_10064	24292	Downstream	NB-ARC domain (NB-ARC)/ Leucine Rich Repeat (LRR_3)
						Ca_10063	2431	Upstream	Phospholipid-transporting ATPase tat-1
	Ca2_16716210	***T/C***	***4.40E-06***	***0.12***	***-1.60***	Ca_22049	1280	Downstream	Protein-serine/threonine phosphatase
						Ca_22050	***13224***	***Upstream***	***Remorin family protein***
	Ca5_12255667	T/C		0.21	-1.34	Ca_17088	9464	Downstream	Anion exchange protein

(Continued)

TABLE 2 Continued

Oligosaccharides	Scaffold position	Allele	*P* value	Minor allele frequency	Allelic effect	Chickpea gene ID	Distance from the SNP (bp)*	SNP position to gene loci	Functional annotation
			4.58E-06			Ca_17087	3021	Upstream	Myb-like DNA-binding protein
	Ca5_42575149	T/A	5.24E-06	0.19	-1.93	Ca_11357	12493	Downstream	F-box only protein 6
						Ca_11356	6617	Upstream	Squamosa promoter-binding-like protein 10-related
	Ca6_44267477	T/C	6.50E-06	0.05	-2.52	Ca_24232	23675	Downstream	Aldo-keto reductase family
						Ca_24233	62720	Upstream	Transposon protein, putative, CACTA, En/Spm sub-class
	Ca4_9041839	G/C	6.76E-06	0.20	-1.86	Ca_08396	21870	Downstream	Thioredoxin-like protein
						Ca_08397	112	upstream	Geraniol 8-hydroxylase
	Ca4_43438450	G/A	7.70E-06	0.06	-2.06	Ca_23689	97	Within the gene	Pollen proteins Ole e I like
	Ca1_41085856	T/A	8.75E-06	0.15	-1.91	Ca_18474	11499	Downstream	Disease resistance protein rpp13-related
						Ca_18475	9169	Upstream	Transposon protein, putative, CACTA, En/Spm sub-class
Stachyose	Ca2_35864450	G/A	2.00E-07	0.12	-7.18	Ca_09762	5676	Within the gene	Armadillo repeat-containing protein 8
	Ca4_6044239	C/T	4.87E-07	0.29	5.04	Ca_03642	543	Within the gene	Syntaxin 16
	Ca8_14016267	C/A	2.07E-06	0.09	-8.60	Ca_22742	9720	Downstream	RNA-binding protein
						Ca_22743	50107	Upstream	Mza15-related
	Ca4_47554909	T/G	5.72E-06	0.10	-7.21	Ca_10834	21050	Downstream	Uncharacterized conserved protein
						Ca_10833	7624	Upstream	Histone deacetylase complex
	Ca2_19532676	T/C	6.21E-06	0.07	-7.19	Ca_25228	65339	Downstream	N-acyl-aliphatic-L-amino acid amidohydrolase
						Ca_24535	57346	Upstream	Ulp1 protease family, C-terminal catalytic domain
	Ca3_23073125	***T/C***	***6.86E-06***	***0.14***	***-5.57***	***Ca_06204***	***14903***	***Downstream***	***Phosphatidylinositol n-acetylglucosaminyltransferase subunit p-like protein***
						Ca_06203	7050	Upstream	U3 small nucleolar RNA-associated protein 20
	Ca4_11202747	T/A	8.39E-06	0.14	-6.06	Ca_04352	12894	Downstream	Small subunit ribosomal protein S1
						Ca_04353	***7468***	***Upstream***	***Type I inositol 1,4,5-trisphosphate 5-phosphatase 1***
Sucrose	Ca8_11491206	C/A	2.70E-06	0.35	-3.97	Ca_16815	6142	Downstream	Transposon protein, CACTA, En/Spm sub-class, expressed
						Ca_16814	6618	Upstream	Prostaglandin-E synthase
	Ca3_11428639	G/A	3.56E-06	0.20	-5.45	Ca_19377	35461	Downstream	Aquaporin tip1-3
						Ca_19378	133058	Upstream	Two-component sensor histidine kinase
	Ca5_14173812	***G/A***	***3.96E-06***	***0.11***	***-5.06***	***Ca_20828***	***30851***	***Downstream***	***Multiple inositol-polyphosphate phosphatase/2,3-bisphosphoglycerate 3-phosphatase***
						Ca_20829	30444	Upstream	Calmodulin-binding protein
	Ca6_2510863	G/A	6.24E-06	0.06	-5.58	Ca_10383	2361	Within the gene	FI03258P
	Ca7_44816569	T/A		0.10	5.91	Ca_15705	21617	Downstream	Zinc knuckle (zf-CCHC)

(Continued)

TABLE 2 Continued

Oligosaccharides	Scaffold position	Allele	*P* value	Minor allele frequency	Allelic effect	Chickpea gene ID	Distance from the SNP (bp)*	SNP position to gene loci	Functional annotation
			7.12E-06			Ca_15706	24202	Upstream	Camp-response element binding protein-related
	Ca3_21765752	C/T	7.67E-06	0.08	5.43	Ca_09529	17032	Downstream	Ethanolamine-phosphate cytidylyltransferase
						Ca_09530	8554	Upstream	GRAS family transcription factor
	Ca1_10947431	T/C	8.62E-06	0.22	3.13	Ca_02666	9994	Downstream	WRKY transcription factor 65-related
						Ca_02665	4707	Upstream	Cyclin-B1-4
	Ca2_33431205	***T/A***	***9.25E-06***	***0.06***	***-6.88***	***Ca_10185***	***2477***	***Downstream***	***Udp-glucose dehydrogenase***
						Ca_10184	44823	Upstream	Lipase containing protein
	Ca6_46242693	T/A	9.41E-06	0.19	-3.29	Ca_13883	34784	Downstream	Unknown
						Ca_13884	20164	Upstream	Phospholipase A(2)/ Phospholipase A2
Total Sugar	Ca7_11316700	A/G	1.34E-06	0.28	-15.00	Ca_09363	20294	Downstream	AhpC/TSA antioxidant enzyme (AhpC-TSA_2)
						Ca_09362	7497	Upstream	CAMP-response element binding protein-related
	Ca4_22983648	T/C	2.57E-06	0.38	18.76	Ca_14460	14194	Downstream	Myb/SANT-like DNA-binding domain
						Ca_14459	67448	Upstream	F26K24.5 protein
	Ca4_17208915	C/T	3.26E-06	0.06	-19.82	Ca_05402	453	Within the gene	V-type H+-transporting ATPase subunit B
	Ca5_1870839	G/A	4.21E-06	0.17	-17.79	Ca_26715	68	Within the gene	DNA damage-binding protein 1
	Ca4_14043315	G/A	4.99E-06	0.37	10.93	Ca_04629	7392	Downstream	Homeobox-leucine zipper protein hdg2
						Ca_04630	2011	Upstream	MKIAA1688 protein
	Ca6_22947128	***A/T***	***5.86E-06***	***0.47***	***15.57***	Ca_11153	5645	Downstream	Peroxisomal targeting signal type 2 receptor
						Ca_11152	***4258***	***Upstream***	***Protein-serine/threonine phosphatase***
	Ca4_30621427	***A/T***	***7.43E-06***	***0.26***	***-9.89***	***Ca_14209***	***8067***	***Downstream***	***UDP-glucose dehydrogenase***
						Ca_14210	47995	Upstream	TBCC domain-containing protein 1
	Ca5_30294813	T/C	9.18E-06	0.36	-10.88	Ca_04706	2728	Downstream	Nipped-b-like protein delangin scc2-related
						Ca_04707	9082	Upstream	Aquaporin transporter

*Distance from the SNP (bp) is calculated as the distance from the SNP location to the start site of the upstream or downstream genes. If the SNP is located within a gene, the distance is calculated as the distance to the start site of the gene.
Probable candidate genes involved in seed oligosaccharides metabolism and transport are bolded and italicized.

sweetening agents. The low RFO lines could be exploited in human and animal food and feed industries. Stachyose and raffinose are considered the most undesirable oligosaccharides present in chickpea. The low stachyose and raffinose accessions ICC 13816 and ICC 6263 can be used in breeding programs as a unique germplasm resource to develop chickpea varieties with improved nutrient utilization and digestibility. The chickpea mini-core collection also shows a considerable amount of variation for ciceritol among accessions. It is believed that these compounds play an important role in protecting plants and seeds against drought stress (Keller and Ludlow, 1993). Ciceritol, a new trisaccharide do not correlate with flatulence, was found high in chickpea accessions reported by Quemener and Brillouet (1983) and Xiaoli et al. (2008). Further research indicated that ciceritol also plays an important role in improving gut health by enhancing the growth of *Lactobacillus, Enterococcus*, and *Bifidobacterium* spp in addition to the production of short-chain fatty acids and is used as a potential source of prebiotics (Zhang et al., 2017). Therefore, increasing ciceritol relatively decreases the flatus potential of chickpea.

10.3389/fpls.2022.1024543

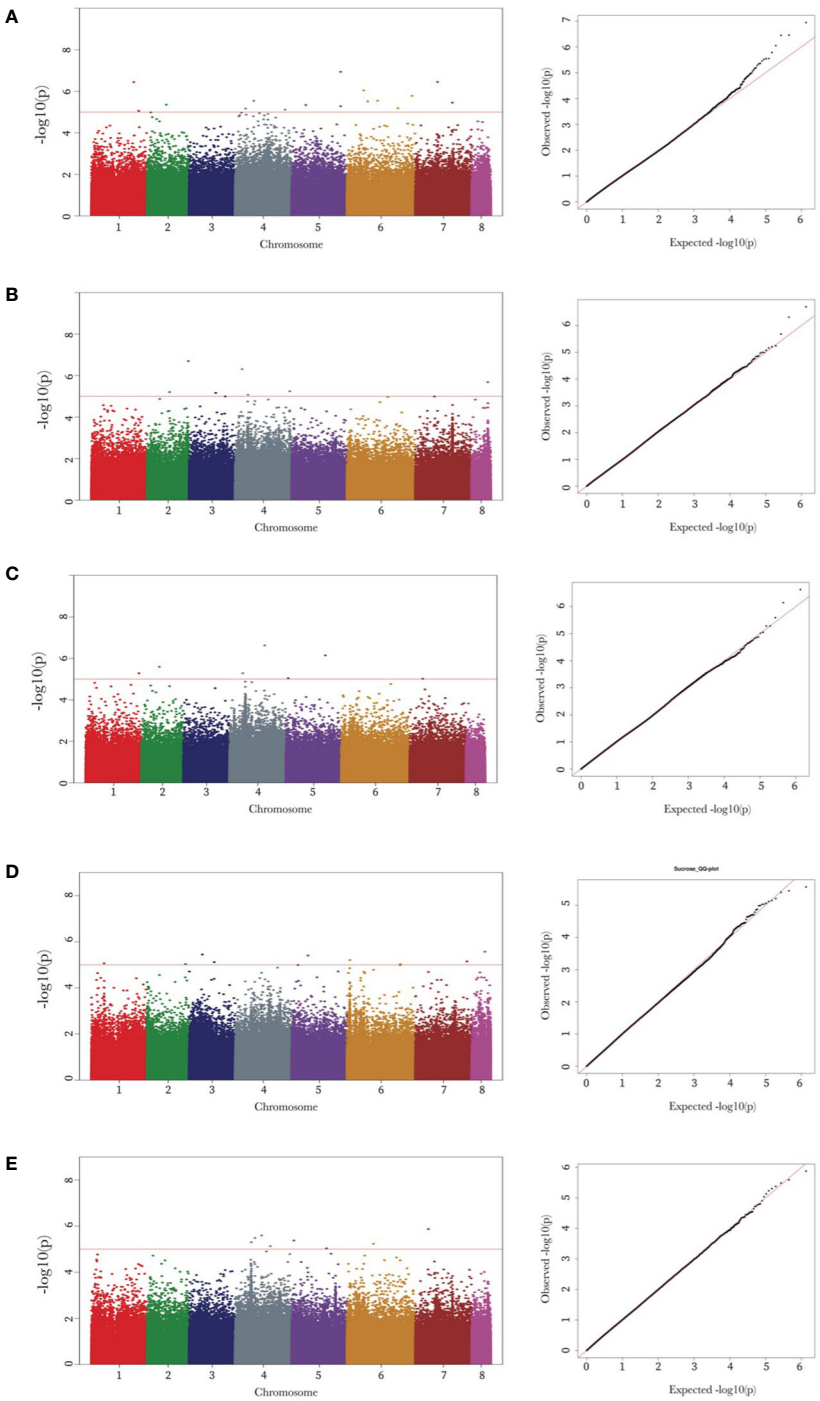


FIGURE 3
Manhattan and quantile-quantile (Q-Q) plots of raffinose, stachyose, ciceritol, sucrose, and total sugars in chickpea. Manhattan and Q-Q plots of the seed oligosaccharides from a to e are as follows: raffinose **(A)**, stachyose **(B)**, ciceritol **(C)**, sucrose **(D)**, and total sugars **(E)**. Negative $\log_{10}$ transformed *P* values (y-axis) are plotted against the physical single nucleotide polymorphism (SNP) position on each chromosome (x-axis). Each circle represents a SNP, and the corresponding SNPs were mentioned trait wise. The dotted red line represents the genome-wide significance threshold as determined by Bonferroni correction at.05. Regions with negative $\log_{10}$ *P* values above the threshold contain quantitative trait loci candidates.

Seed oligosaccharide candidates

There are four genes - two associated with stachyose (*Ca_06204: phosphatidylinositol N-acetylglucosaminyltransferase; Ca_04353: Type I inositol 1,4,5-trisphosphate 5-phosphatase*) and two associated with sucrose (*Ca_20828: inositol-polyphosphate phosphatase; Ca_10185:UDP-glucose dehydrogenase*) – have been recognized as indispensably involved in various signaling pathways in plants *via* mediating the phospholipidation and phosphatization of *Myo-inositol* and its derivatives like sucrose and RFOs. Studies reported the crosstalk linkage between inositol signaling and sugar metabolism in plants (Saddhe et al., 2021; Lou et al., 2007; Yang et al., 2007). In 2008, Ananieva et al. (2008) discovered the *in-vitro* interactions between the *myo-inositol polyphosphate 5-phosphatase (5PTase13)* and the *sucrose nonfermenting-1-related kinase (SnRK1.1)* and identified *5PTase13* regulated *SnRK1* activity under different sugar conditions. Plant *SnRK1* is known for its key role at the interface among sugar metabolism, stress signaling, and other physiological developmental processes like seed germination and seedling growth (Radchuk et al., 2006; Baena-González et al., 2007; Jossier et al., 2009; Hulsmans et al., 2016; Elango et al., 2022b). *SnRK1* has been identified in higher plants and two other subfamilies – *SnRK2* and *SnRK3* (Halford and Hey, 2009). Purcell et al. (1998) demonstrated that *SnRK1* plays an essential role in the regulation of *Suc synthase* expression in potatoes (*Solanum tuberosum*), and later Tiessen et al. (2003) recognized that *SnRK1* also regulates starch biosynthesis. The same finding about *SnRK1*'s role involved in starch synthesis was also identified in pollen grains of barley (*Hordeum vulgare*) (Zhang et al., 2001). Our findings indicate that in chickpea, the stachyose and sucrose biosynthesis is mediated by various kinds of *inositol-polyphosphate phosphatase*, potentially through their regulation of *SnRK* families. Additionally, the genes involved in the RFO biosynthetic pathway have been identified in soybean (*Glycine max*) and common bean (*Phaseolus vulgaris*). de Koning et al. (2021) identified three *galactinol synthase* (*GolS*) genes in common bean, named *PvGolS1, PvGolS2*, and *PvGolS3*. *GolS* crosslinks between *inositol* and RFO biosynthesis, and *GolS* are the primary checkpoint for RFO biosynthesis *via inositols* (Sengupta et al., 2015).

We also found that three candidate genes encode for *remorin* protein – two genes (*Ca_17399* and *Ca_22050*) are associated with raffinose content, and the other gene (*Ca_11152*) is linked with total sugar content in chickpea seeds. RFOs serve as a major transport form of carbohydrates in the vascular system in plants (Ayre et al., 2003; Johnson et al., 2020; Ren et al., 2021). *Remorin* is a kind of plant-specific membrane-bound protein and has been identified in *Arabidopsis thaliana*, *Nicotiana tabacum*, *Medicago truncatula*, and *Lycopersicon esculentum* (Watson et al., 2003; Bariola et al., 2004; Marmagne et al., 2004; Mongrand et al., 2004; Sazuka et al., 2004; Nelson et al., 2006; Valot et al., 2006). In Arabidopsis, 16 genes have been identified in the *REM* family, and among them, *REM1.2, REM1.3, and REM1.4* from the *REM1* subfamily were found to exist ubiquitously in the majority of the tissues (Huang et al., 2019). Previous studies demonstrated that *remorin* proteins are localized in the plasma membrane and plasmodesmata of phloem companion cells and regulate photoassimilate translocation *via* reducing plasmodesmata permeability in the symplastic system in rice (*Oryza sativa*) (Gui et al., 2014). As an example, in rice, over-expressed *remorin* gene *gsd1-D* in the dominant mutant (*grain setting defect1-Dominant*) showed a grain setting-deficient phenotype of reduced grain setting rate, reversible accumulation of carbohydrate in leaves, and reduced synthesis of soluble sugar concentration in phloem exudates (Gui et al., 2014).

Three candidate genes are also identified as *UDP-glucose dehydrogenase (UGD)* – gene *Ca_14209* is simultaneously associated with two traits (ciceritol and total sugar contents). The other two genes are *Ca_27229* (associated with raffinose content) and *Ca_10185* (associated with sucrose content). *UGD* is a key enzyme in carbohydrate metabolism and has been identified in soybean (*Glycine max*), maize (*Zea mays* L.), cotton (*Gossypium hirsutum*), and Arabidopsis (*Arabidopsis thaliana*) (Kärkönen et al., 2005; Kohlberger et al., 2018; Jia et al., 2021). The overexpression of UDG-coding gene *PeUGDH4* in Arabidopsis leads to a significant increase in hemicellulose synthesis (Yang et al., 2020), indicating its critical role in plant cell wall synthesis. *UGD* converts UDP-glucose to UDP-glucuronic acid, providing the precursor for hemicellulose and pectin biosynthesis - the two confound components in the primary cell wall matrix (Oka and Jigami, 2006). And later, *UGD* has also been identified to be highly involved in the secondary cell wall construction in Moso bamboo (*Phyllostachys edulis*) (Yang et al., 2020). Besides being incorporated into the cell wall, the remainder forms of carbohydrates can be small molecule oligosaccharides such as RFOs. The RFOs biosynthesis pathways started with UDP-galactose being converted to UDP-galacturonic acid (*UDP-GalA*) by UG4E UDP-glucose 4′-epimerase; or alternatively UDP-glucose being converted to myo-inositol (Karner et al., 2004; Joët et al., 2009). Then both *UDP-GAL* and *myo-inositol* can be the precursors of the galactinol biosynthesis by *galactinol synthase GolS* (Keller and Pharr, 1996). Galactinol is believed to be the only known galactosyl donor to RFOs (Sprenger and Keller, 2000). In summary, the cell wall polysaccharide (CWP) and RFOs biosynthesis pathways are interconnected but also competitive for the upstream precursor UDP-glucose. The increase in UGD activity could lead to the increased production of CWP; however, at the same time, it could diminish RFOs and other carbohydrates production.

Conclusion

The present study identified potential candidate genes regulating the biosynthesis and transport of seed oligosaccharides in chickpea. We have identified 48 SNPs associated with five sugar types. Nine genes (*Ca_06204, Ca_04353*, and *Ca_20828*: *Phosphatidylinositol N-acetylglucosaminyltransferase*; *Ca_17399* and *Ca_22050*: *Remorin proteins*; *Ca_11152*: *Protein-serine/threonine phosphatase; Ca_10185, Ca_14209*, and *Ca_27229*: *UDP-glucose dehydrogenase*) were identified as potential candidate genes for sugar metabolism and transport in chickpea. The accessions with low RFOs and high sucrose contents may be utilized in breeding specialty chickpeas. The identified candidates could be exploited in marker-assisted breeding, genomic selection, and genetic engineering to improve the sugar profiles in legumes and other crop species.

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

DE- designed experiments, collected and analyzed data, and wrote the manuscript. WW- analyzed data and wrote the manuscript. MT- analyzed the data. SS- designed experiments, collected and analyzed data. BR- analyzed and wrote the manuscript. RV- edited the manuscript. All authors contributed to the article and approved the submitted version.

Funding

The open access publication fees for this article were covered by the Iowa State University Library.

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.1024543/full#supplementary-material

References

Aguilera, Y., Martín-Cabrejas, M. A., Benítez, V., Mollá, E., López-Andréu, F. J., and Esteban, R. M. (2009). Changes in carbohydrate fraction during dehydration process of common legumes. *J. Food Composition Anal.* 22, 678–683. doi: 10.1016/j.jfca.2009.02.012

Ananieva, E. A., Gillaspy, G. E., Ely, A., Burnette, R. N., and Erickson, F.l. (2008). Interaction of the WD40 domain of a myoinositol polyphosphate 5-phosphatase with SnRK1 links inositol, sugar, and stress signaling. *Plant Physiol.* 148, 1868–1882. doi: 10.1104/PP.108.130575

Ayre, B. G., Keller, F., and Turgeon, R. (2003). Symplastic continuity between companion cells and the translocation stream: Long-distance transport is controlled by retention and retrieval mechanisms in the phloem. *Plant Physiol.* 131, 1518–1528. doi: 10.1104/PP.012054

Baena-González, E., Rolland, F., Thevelein, J. M., and Sheen, J. (2007). A central integrator of transcription networks in plant stress and energy signalling. *Nature* 448, 938–942. doi: 10.1038/NATURE06069

Bariola, P. A., Retelska, D., Stasiak, A., Kammerer, R. A., Fleming, A., Hijri, M., et al. (2004). Remorins form a novel family of coiled coil-forming oligomeric and filamentous proteins associated with apical, vascular and embryonic tissues in plants. *Plant Mol. Biol.* 55, 579–594. doi: 10.1007/S11103-004-1520-4

Blackman, S. A., Obendorf, R. L., and Leopold, A. C. (1992). Maturation proteins and sugars in desiccation tolerance of developing soybean seeds. *Plant Physiol.* 100, 225–230. doi: 10.1104/pp.100.1.225

Calloway, D. H., and Murphy, E. L. (1968). The use of expired air to measure intestinal gas formation. *Ann. N Y Acad. Sci.* 150, 82–95. doi: 10.1111/J.1749-6632.1968.TB19034.X

Chibbar, R. N., Ambigaipalan, P., and Hoover, R. (2010). REVIEW: Molecular diversity in pulse seed starch and complex carbohydrates and its role in human nutrition and health. *Cereal Chem.* 87, 342–352. doi: 10.1094/CCHEM-87-4-0342

de Koning, R., Kiekens, R., Toili, M. E. M., and Angenon, G. (2021). Identification and expression analysis of the genes involved in the raffinose family oligosaccharides pathway of phaseolus vulgaris and glycine max. *Plants* 10, 1–21. doi: 10.3390/plants10071465

de Las Mercedes Dana, M., Pintor-Toro, J. A., and Cubero, B. (2006). Transgenic tobacco plants overexpressing chitinases of fungal origin show enhanced resistance to biotic and abiotic stress agents. *Plant Physiol.* 142, 722. doi: 10.1104/PP.106.086140

Delumen, B. O. (1992). Molecular strategies to improve protein-quality and reduce flatulence in legumes - a review. *Food Structure* 11, 33–46.

Dierking, E. C., and Bilyeu, K. D. (2008). Association of a soybean raffinose synthase gene with low raffinose and stachyose seed phenotype. *Plant Genome* 1, 135–145. doi: 10.3835/plantgenome2008.06.0321

Dobrenel, T., Marchive, C., Azzopardi, M., Clément, G., Moreau, M., Sormani, R., et al. (2013). Sugar metabolism and the plant target of rapamycin kinase: a sweet operaTOR? *Front. Plant Sci.* 4. doi: 10.3389/fpls.2013.00093

Egbe, I. A., and Akinyele, I. O. (1990). Effect of cooking on the antinutritional factors of lima beans (Phaseolus lunatus). *Food Chem.* 35, 81–87. doi: 10.1016/0308-8146(90)90022-V

Elango, D., Rajendran, K., van der Laan, L., Sebastiar, S., Raigne, J., Thaiparambil, N. A., et al. (2022a). Raffinose family oligosaccharides: Friend or foe for human and plant health? *Front. Plant Sci.* 13. doi: 10.3389/FPLS.2022.829118/BIBTEX

Elango, D., Wang, X., Bhatnagar, R. S., Tan, Q., Gaffoor, I., Hu, Z., et al. (2022b). Association genetics of early season cold and late season frost tolerance in sorghum bicolor. *Crop Sci.* 62, 1844–1865. doi: 10.1002/csc2.20810

Gangl, R., and Tenhaken, R. (2016). Raffinose Family Oligosaccharides Act As Galactose Stores in Seeds and Are Required for Rapid Germination of Arabidopsis in the Dark. *Front Plant Sci* 7, 1–15. doi: 10.3389/fpls.2016.01115

Gangola, M. P., Jaiswal, S., Kannan, U., Gaur, P. M., Båga, M., and Chibbar, R. N. (2016). Galactinol synthase enzyme activity influences raffinose family oligosaccharides (RFO) accumulation in developing chickpea (Cicer arietinum l.) seeds. *Phytochemistry* 125, 88–98. doi: 10.1016/j.phytochem.2016.02.009

Graham, P. H., and Vance, C. P. (2003). Legumes: Importance and constraints to greater use. *Plant Physiol.* 131, 872–877. doi: 10.1104/PP.017004

Gui, J., Liu, C., Shen, J., and Li, L. (2014). Grain setting defect1, encoding a remorin protein, affects the grain setting in rice through regulating plasmodesmatal conductance. *Plant Physiol.* 166, 1463–1478. doi: 10.1104/PP.114.246769

Gulewicz, P., Szymaniec, S., Bubak, B., Frias, J., Vidal-Valverde, C., Trojanowska, K., et al. (2002). Biological activity of α-galactoside preparations from lupinus angustifolius l. and pisum sativum l. seeds. *J. Agric. Food Chem.* 50, 384–389. doi: 10.1021/jf010973y

Halford, N. G., and Hey, S. J. (2009). Snf1-related protein kinases (SnRKs) act within an intricate network that links metabolic and stress signalling in plants. *Biochem. J.* 419, 247–259. doi: 10.1042/BJ20082408

Han, I. H., and Baik, B. K. (2006). Oligosaccharide content and composition of legumes and their reduction by soaking, cooking, ultrasound, and high hydrostatic pressure. *Cereal Chem.* 83, 428–433. doi: 10.1094/CC-83-0428

Huang, D., Sun, Y., Ma, Z., Ke, M., Cui, Y., Chen, Z., et al. (2019). Salicylic acid-mediated plasmodesmal closure *via* remorin-dependent lipid organization. *Proc. Natl. Acad. Sci. U.S.A.* 116, 21274–21284. doi: 10.1073/PNAS.1911892116

Hulsmans, S., Rodriguez, M., de Coninck, B., and Rolland, F. (2016). The SnRK1 energy sensor in plant biotic interactions. *Trends Plant Sci.* 21, 648–661. doi: 10.1016/J.TPLANTS.2016.04.008/ATTACHMENT/3B97F173-DE7E-48F8-83F5-FF0FBC3B99C5/MMC1.MP4

Jia, T., Ge, Q., Zhang, S., Zhang, Z., Liu, A., Fan, S., et al. (2021). UDP-Glucose dehydrogenases: Identification, expression, and function analyses in upland cotton (Gossypium hirsutum). *Front. Genet.* 111648. doi: 10.3389/FGENE.2020.597890/BIBTEX

Joët, T., Laffargue, A., Salmona, J., Doulbeau, S., Descroix, F., Bertrand, B., et al. (2009). Metabolic pathways in tropical dicotyledonous albuminous seeds: Coffea arabica as a case study. *New Phytol.* 182, 146–162. doi: 10.1111/J.1469-8137.2008.02742.X

Johnson, N., Johnson, C. R., Thavarajah, P., Kumar, S., and Thavarajah, D. (2020). The roles and potential of lentil prebiotic carbohydrates in human and plant health. *Plants People Planet* 2, 310–319. doi: 10.1002/PPP3.10103

Jood, S., Mehta, U., Bhat, C. M., and Singh, R. (1985). Effect of processing on flatus-producing factors in legumes. *J. Agric. Food Chem.* 33, 268–271. doi: 10.1021/jf00062a028

Jossier, M., Bouly, J. P., Meimoun, P., Arjmand, A., Lessard, P., Hawley, S., et al. (2009). SnRK1 (SNF1-related kinase 1) has a central role in sugar and ABA signalling in arabidopsis thaliana. *Plant J.* 59, 316–328. doi: 10.1111/J.1365-313X.2009.03871.X

Jukanti, A. K., Gaur, P. M., Gowda, C. L. L., and Chibbar, R. N. (2012). Nutritional quality and health benefits of chickpea (Cicer arietinum l.): a review. *Br. J. Nutr.* 108, S11–S26. doi: 10.1017/S0007114512000797

Kärkönen, A., Murigneux, A., Martinant, J. P., Pepey, E., Tatout, C., Dudley, B. J., et al. (2005). UDP-Glucose dehydrogenases of maize: a role in cell wall pentose biosynthesis. *Biochem. J.* 391, 409–415. doi: 10.1042/BJ20050800

Karlsson, M., and Stenlid, J. (2008). Comparative evolutionary histories of the fungal chitinase gene family reveal non-random size expansions and contractions due to adaptive natural selection. *Evol. Bioinf.* 2008, 47–60. doi: 10.4137/ebo.s604

Karner, U., Peterbauer, T., Raboy, V., Jones, D. A., Hedley, C. L., and Richter, A. (2004). Myo-inositol and sucrose concentrations affect the accumulation of raffinose family oligosaccharides in seeds. *J. Exp. Bot.* 55, 1981–1987. doi: 10.1093/JXB/ERH216

Keller, F. (1992). Transport of stachyose and sucrose by vacuoles of Japanese artichoke (Stachys sieboldii) tubers. *Plant Physiol.* 98, 442–445. doi: 10.1104/PP.98.2.442

Keller, F., and Ludlow, M. M. (1993). Carbohydrate metabolism in drought-stressed leaves of pigeonpea (Cajanus cajan). *J. Exp. Bot.* 44, 1351–1359. doi: 10.1093/jxb/44.8.1351

Keller, F., and Pharr, D. M. (1996). "Metabolism of carbohydrates in sinks and sources: galactosyl-sucrose oligosaccharides," in *Photoassimilate distribution in plants and crops*. Eds. E. Zamski and A. A. Schaffer (New York: Taylor & Francis), 157–183.

Kohlberger, M., Thalhamer, T., Weiss, R., and Tenhaken, R. (2018). Arabidopsis MAP-kinase 3 phosphorylates UDP-glucose dehydrogenase: a key enzyme providing UDP-sugar for cell wall biosynthesis. *Plant Mol. Biol. Rep.* 36, 870. doi: 10.1007/S11105-018-1130-Y

Lev-Yadun, S., Gopher, A., and Abbo, S. (2000). The cradle of agriculture. *Sci. (1979)* 288, 1602–1603. doi: 10.1126/science.288.5471.1602

Liu, X., Huang, M., Fan, B., Buckler, E. S., and Zhang, Z. (2016). Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. *PloS Genet.* 12, e1005767. doi: 10.1371/journal.pgen.1005767

Lou, Y., Gou, J. Y., and Xue, H. W. (2007). PIP5K9, an arabidopsis phosphatidylinositol monophosphate kinase, interacts with a cytosolic invertase to negatively regulate sugar-mediated root growth. *Plant Cell* 19, 163–181. doi: 10.1105/TPC.106.045658

Marmagne, A., Roue, M. A., Ferro, M., Rolland, N., Alcon, C., Joyard, J., et al. (2004). Identification of new intrinsic proteins in arabidopsis plasma membrane proteome. *Mol. Cell. Proteomics* 3, 675–691. doi: 10.1074/MCP.M400001-MCP200

Mongrand, S., Morel, J., Laroche, J., Claverol, S., Carde, J. P., Hartmann, M. A., et al. (2004). Lipid rafts in higher plant cells: purification and characterization of triton X-100-insoluble microdomains from tobacco plasma membrane. *J. Biol. Chem.* 279, 36277–36286. doi: 10.1074/JBC.M403440200

Nelson, C. J., Hegeman, A. D., Harms, A. C., and Sussman, M. R. (2006). A quantitative analysis of arabidopsis plasma membrane using trypsin-catalyzed 18O labeling. *Mol. Cell. Proteomics* 5, 1382–1395. doi: 10.1074/MCP.M500414-MCP200/ATTACHMENT/F5709B47-DBEB-49EE-9EE7-676C9E281C9C/MMC1.ZIP

Nishizawa-Yokoi, A., Yabuta, Y., and Shigeoka, S. (2008). The contribution of carbohydrates including raffinose family oligosaccharides and sugar alcohols to protection of plant cells from oxidative damage. *Plant Signal Behav.* 3, 1016–1018. doi: 10.4161/psb.6738

Obendorf, R. L., and Górecki, R. J. (2012). Soluble carbohydrates in legume seeds. *Seed Sci. Res.* 22, 219–242. doi: 10.1017/S0960258512000104

Oka, T., and Jigami, Y. (2006). Reconstruction of *de novo* pathway for synthesis of UDP-glucuronic acid and UDP-xylose from intrinsic UDP-glucose in saccharomyces cerevisiae. *FEBS J.* 273, 2645–2657. doi: 10.1111/J.1742-4658.2006.05281.X

Peterbauer, T., Karner, U., Mucha, J., Mach, L., Jones, D. A., Hedley, C. L., et al. (2003). Enzymatic control of the accumulation of verbascose in pea seeds. *Plant Cell Environ.* 26, 1385–1391. doi: 10.1046/j.0016-8025.2003.01063.x

Peterbauer, T., Lahuta, L. B., Blöchl, A., Mucha, J., Jones, D. A., Hedley, C. L., et al. (2001). Analysis of the raffinose family oligosaccharide pathway in pea seeds with contrasting carbohydrate composition. *Plant Physiol.* 127, 1764–1772. doi: 10.1104/pp.010534

Purcell, P. C., Smith, A. M., and Halford, N. G. (1998). Antisense expression of a sucrose non-fermenting-1-related protein kinase sequence in potato results in decreased expression of sucrose synthase in tubers and loss of sucrose-inducibility of sucrose synthase transcripts in leaves. *Plant J.* 14, 195–202. doi: 10.1046/J.1365-313X.1998.00108.X

Qiu, D., Vuong, T., Valliyodan, B., Shi, H., Guo, B., Shannon, J. G., et al. (2015). Identification and characterization of a stachyose synthase gene controlling reduced stachyose content in soybean. *Theor. Appl. Genet.* 128, 2167–2176. doi: 10.1007/S00122-015-2575-0

Quemener, B., and Brillouet, J. M. (1983). Ciceritol, a pinitol digalactoside form seeds of chickpea, lentil and white lupin. *Phytochemistry* 22, 1745–1751. doi: 10.1016/S0031-9422(00)80263-0

Radchuk, R., Radchuk, V., Weschke, W., Borisjuk, L., and Weber, H. (2006). Repressing the expression of the SUCROSE NONFERMENTING-1-RELATED PROTEIN KINASE gene in pea embryo causes pleiotropic defects of maturation similar to an abscisic acid-insensitive phenotype. *Plant Physiol.* 140, 263. doi: 10.1104/PP.105.071167

Raja, R. B., Balraj, R., Agasimani, S., Dinakaran, E., Thiruvengadam, V., Kannan Bapu, J. R., et al. (2015). Determination of oligosaccharide fraction in a worldwide germplasm collection of chickpea ('Cicer arietinum'L.) using high performance liquid chromatography. *Aust. J. Crop Sci.* 9, 605–613.

Ren, Y., Li, M., Guo, S., Sun, H., Zhao, J., Zhang, J., et al. (2021). Evolutionary gain of oligosaccharide hydrolysis and sugar transport enhanced carbohydrate partitioning in sweet watermelon fruits. *Plant Cell* 33, 1554–1573. doi: 10.1093/PLCELL/KOAB055

Saddhe, A. A., Manuka, R., and Penna, S. (2021). Plant sugars: Homeostasis and transport under abiotic stress in plants. *Physiol Plant* 171, 739–755. doi: 10.1111/PPL.13283

Sazuka, T., Keta, S., Shiratake, K., Yamaki, S., and Shibata, D. (2004). A proteomic approach to identification of transmembrane proteins and membrane-anchored proteins of arabidopsis thaliana by peptide sequencing. *DNA Res.* 11, 101–113. doi: 10.1093/DNARES/11.2.101

Sengupta, S., Mukherjee, S., Basak, P., and Majumder, A. L. (2015). Significance of galactinol and raffinose family oligosaccharide synthesis in plants. *Front. Plant Sci.* 6. doi: 10.3389/fpls.2015.00656

Singh, U. (1985). Nutritional quality of chickpea (Cicer arietinum l.): current status and future research needs. *Plant Foods Hum. Nutr.* 35, 339–351. doi: 10.1007/BF01091779

Sosulski, F. W., Elkowicz, L., and Reichert, R. D. (1982). Oligosaccharides in eleven legumes and their air-classified protein and starch fractions. *J. Food Sci.* 47, 498–502. doi: 10.1111/j.1365-2621.1982.tb10111.x

Souframanien, J., Roja, G., and Gopalakrishna, T. (2014). Genetic variation in raffinose family oligosaccharides and sucrose content in blackgram [Vigna mungo l. (Hepper)]. *J. Food Legumes* 27, 37–41.

Sprenger, N., and Keller, F. (2000). Allocation of raffinose family oligosaccharides to transport and storage pools in ajuga reptans: the roles of two distinct galactinol synthases. *Plant J.* 21, 249–258. doi: 10.1046/J.1365-313X.2000.00671.X

Su, H., Chen, J., Miao, S., Deng, K., Liu, J., Zeng, S., et al. (2019). Lotus seed oligosaccharides at various dosages with prebiotic activity regulate gut microbiota and relieve constipation in mice. *Food and Chemical Toxicology* 134, 1–12. doi: 10.1016/j.fct.2019.110838

Tahir, M., Lindeboom, N., Båga, M., Vandenberg, A., and Chibbar, R. (2011). Composition and correlation between major seed constituents in selected lentil (Lens culinaris medik) genotypes. *Can. J. Plant Sci.* 91, 825–835. doi: 10.4141/cjps2011-010

Taji, T., Ohsumi, C., Iuchi, S., Seki, M., Kasuga, M., Kobayashi, M., et al. (2002). Important roles of drought- and cold-inducible genes for galactinol synthase in stress tolerance in arabidopsis thaliana. *Plant J.* 29, 417–426. doi: 10.1046/j.0960-7412.2001.01227.x

Tiessen, A., Prescha, K., Branscheid, A., Palacios, N., McKibbin, R., Halford, N. G., et al. (2003). Evidence that SNF1-related kinase and hexokinase are involved in separate sugar-signalling pathways modulating post-translational redox activation of ADP-glucose pyrophosphorylase in potato tubers. *Plant J.* 35, 490–500. doi: 10.1046/J.1365-313X.2003.01823.X

Unda, F., Kim, H., Hefer, C., Ralph, J., and Mansfield, S. D. (2017). Altering carbon allocation in hybrid poplar (Populus alba × grandidentata) impacts cell wall growth and development. *Plant Biotechnol. J.* 15, 865–878. doi: 10.1111/pbi.12682

Upadhyaya, H. D., and Ortiz, R. (2001). A mini core subset for capturing diversity and promoting utilization of chickpea genetic resources in crop improvement. *Theor. Appl. Genet.* 102, 1292–1298. doi: 10.1007/s00122-001-0556-y

Valot, B., Negroni, L., Zivy, M., Gianinazzi, S., and Dumas-Gaudot, E. (2006). A mass spectrometric approach to identify arbuscular mycorrhiza-related proteins in root plasma membrane fractions. *Proteomics* 6 (Suppl 1), S145–S155. doi: 10.1002/PMIC.200500403

Van den Ende, W. (2013). Multifunctional fructans and raffinose family oligosaccharides. *Front. Plant Sci.* 4. doi: 10.3389/fpls.2013.00247

Varshney, R. K., Song, C., Saxena, R. K., Azam, S., Yu, S., Sharpe, A. G., et al. (2013). Draft genome sequence of chickpea (Cicer arietinum) provides a resource for trait improvement. *Nat. Biotechnol.* 31, 240–246. doi: 10.1038/nbt.2491

Varshney, R. K., Thudi, M., Roorkiwal, M., He, W., Upadhyaya, H. D., Yang, W., et al. (2019). Resequencing of 429 chickpea accessions from 45 countries provides insights into genome diversity, domestication and agronomic traits. *Nat. Genet.* 51, 857–864. doi: 10.1038/s41588-019-0401-3

Wang, J., and Zhang, Z. (2021). GAPIT version 3: Boosting power and accuracy for genomic association and prediction. *Genomics Proteomics Bioinf.* 19, 629–640. doi: 10.1016/j.gpb.2021.08.005

Watson, B. S., Asirvatham, V. S., Wang, L., and Sumner, L. W. (2003). Mapping the proteome of barrel medic (Medicago truncatula). *Plant Physiol.* 131, 1104–1123. doi: 10.1104/PP.102.019034

Xiaoli, X., Liyi, Y., Shuang, H., Wei, L., Yi, S., Hao, M., et al. (2008). Determination of oligosaccharide contents in 19 cultivars of chickpea (Cicer arietinum l) seeds by high performance liquid chromatography. *Food Chem.* 111, 215–219. doi: 10.1016/J.FOODCHEM.2008.03.039

Xin, Y., Wang, D., Han, S., Li, S., Gong, N., Fan, Y., et al. (2021). Characterization of the chitinase gene family in mulberry (Morus notabilis) and MnChi18 involved in resistance to botrytis cinerea. *Genes (Basel)* 13, 1–12. doi: 10.3390/GENES13010098/S1

Yan, S., Liu, Q., Li, W., Yan, J., and Fernie, A. R. (2022). Raffinose Family Oligosaccharides: Crucial Regulators of Plant Development and Stress Responses. *CRC Crit Rev Plant Sci* 41, 286–303. doi: 10.1080/07352689.2022.2111756

Yang, Y., Kang, L., Wu, R., Chen, Y., and Lu, C. (2020). Genome-wide identification and characterization of UDP-glucose dehydrogenase family genes in moso bamboo and functional analysis of PeUGDH4 in hemicellulose synthesis. *Sci. Rep. 2020 10:1* 10, 1–13. doi: 10.1038/s41598-020-67227-8

Yang, J. I., Perera, I. Y., Brglez, I., Davis, A. J., Stevenson-Paulik, J., Phillippy, B. Q., et al. (2007). Increasing plasma membrane Phosphatidylinositol(4,5) Bisphosphate biosynthesis increases phosphoinositide metabolism in nicotiana tabacum. *Plant Cell* 19, 1603–1616. doi: 10.1105/TPC.107.051367

Zhang, Y. J., Ren, L. L., Lin, X. Y., Han, X. M., Wang, W., and Yang, Z. L. (2022). Molecular evolution and functional characterization of chitinase gene family in populus trichocarpa. *Gene* 822, 1–10. doi: 10.1016/J.GENE.2022.146329

Zhang, Y., Shewry, P. R., Jones, H., Barcelo, P., Lazzeri, P. A., and Halford, N. G. (2001). Expression of antisense SnRK1 protein kinase sequence causes abnormal pollen development and male sterility in transgenic barley. *Plant J.* 28, 431–441. doi: 10.1046/J.1365-313X.2001.01167.X

Zhang, Y., Su, D., He, J., Dai, Z., Asad, R., Ou, S., et al. (2017). Effects of ciceritol from chickpeas on human colonic microflora and the production of short chain fatty acids by in vitro fermentation. *LWT - Food Sci. Technol.* 79, 294–299. doi: 10.1016/J.LWT.2017.01.040

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 03 November 2022
DOI 10.3389/fpls.2022.1050462

OPEN ACCESS

EDITED BY
Mingyue Gou,
Henan Agricultural University, China

REVIEWED BY
Beth Allyn Krizek,
University of South Carolina, United States
Cédric Finet,
National University of Singapore, Singapore

*CORRESPONDENCE
Lu Han
lhan@sdu.edu.cn

SPECIALTY SECTION
This article was submitted to Functional and Applied Plant Genomics, a section of the journal Frontiers in Plant Science

RECEIVED 21 September 2022
ACCEPTED 20 October 2022
PUBLISHED 03 November 2022

CITATION
Wang X, Zhang J, Zhang J, Zhou C and Han L (2022) Genome-wide characterization of *AINTEGUMENTA-LIKE* family in *Medicago truncatula* reveals the significant roles of *AINTEGUMENTAs* in leaf growth. *Front. Plant Sci.* 13:1050462. doi: 10.3389/fpls.2022.1050462

Genome-wide characterization of *AINTEGUMENTA-LIKE* family in *Medicago truncatula* reveals the significant roles of *AINTEGUMENTAs* in leaf growth

Xiao Wang, Juanjuan Zhang, Jing Zhang, Chuanen Zhou and Lu Han*

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

AINTEGUMENTA-LIKE (AIL) transcription factors are widely studied and play crucial roles in plant growth and development. However, the functions of the *AIL* family in legume species are largely unknown. In this study, 11 *MtAIL* genes were identified in the model legume *Medicago truncatula*, of which four of them are *MtANTs*. *In situ* analysis showed that *MtANT1* was highly expressed in the shoot apical meristem (SAM) and leaf primordium. Characterization of *mtant1 mtant2 mtant3 mtant4* quadruple mutants and *MtANT1*-overexpressing plants revealed that *MtANTs* were not only necessary but also sufficient for the regulation of leaf size, and indicated that they mainly function in the regulation of cell proliferation during secondary morphogenesis of leaves in *M. truncatula*. This study systematically analyzed the *MtAIL* family at the genome-wide level and revealed the functions of *MtANTs* in leaf growth. Thus, these genes may provide a potential application for promoting the biomass of legume forages.

KEYWORDS

***Medicago truncatula*, *AILs*, *ANT*, leaf size, leaf morphogenesis**

Introduction

The AINTEGUMENTA-LIKE (AIL) transcription factors have been widely studied in plants (Klucher et al., 1996; Mizumoto et al., 2009; Karlberg et al., 2011; Li and Xue, 2011; Rigal et al., 2012; Horstman et al., 2014; Kuluev et al., 2015; Bui et al., 2017; Ding et al., 2018; Zhao et al., 2019b; Liu et al., 2020; Shen et al., 2020; Miao et al., 2021; Han et al., 2022). They belong to the APETALA 2 (AP2)-like subfamily, which is characterized by two putative DNA-binding AP2 domains and one conserved linker region (Kim et al., 2006). The AP2-like subfamily can be divided into three groups: euAP2, basalANT, and

euANT (Kim et al., 2006). Unlike *euAP2* genes, the *basalANT* and *euANT* genes can not be recognized by *miR172* and are differentiated by specific sequence signatures (Kim et al., 2006; Dipp-Álvarez and Cruz-Ramírez, 2019).

In *Arabidopsis thaliana*, the euANT proteins are also known as AILs and consist of *ANT*, *AIL1*, *AIL2*/*BBM*/*PLT4*, *AIL3*/*PLT1*, *AIL4*/*PLT2*, *AIL5*/*PLT5*, *AIL6*/*PLT3* and *AIL7*/*PLT7* (Nole-Wilson et al., 2005; Horstman et al., 2014). *ANT* mainly regulates cell division and cell differentiation in leaves and floral organs. For instance, *ANT* regulates the size of lateral organs by controlling cell proliferation during organogenesis (Mizukami and Fischer, 2000). *ANT* also participates in the establishment of adaxial-abaxial polarity. *ANT* acts with the abaxial-specifying gene *FILAMENTOUS FLOWER* (*FIL*) to up-regulate the expression of the adaxial gene *PHABULOSA* (Nole-Wilson and Krizek, 2006). Meanwhile, *ANT* and *AIL5*/*6*/*7* function partially redundantly in flower development, and the direct targets of *ANT* and *AIL6* include *LEAFY* and other genes involved in polarity establishment, meristem and flower development, as well as auxin signaling pathway (Krizek, 2009; Krizek, 2015; Yamaguchi et al., 2016; Krizek et al., 2020; Krizek et al., 2021). *ANT*, *AIL6*, and *AIL7* regulate SAM function; the SAM of the *ant ail6 ail7* triple mutant terminates after the production of only a few leaves (Mudunkothge and Krizek, 2012). Furthermore, *BBM*/*AIL2*, *PLT1*/*AIL3*, *PLT2*/*AIL4* and *PLT3*/*AIL6* regulate cell proliferation during embryogenesis and root apical meristem maintenance (Horstman et al., 2014). These genes function in stem cell identity in the root and promote cell division of the stem cell daughters (Galinha et al., 2007). *PLT1*/*AIL3* and *PLT2*/*AIL4* also play essential roles in the quiescent center specification and stem cells maintenance (Aida et al., 2004). *BBM*/*AIL2* is involved in the embryo and endosperm development, and ectopic expression of *BBM*/*AIL2* leads to the production of somatic embryos on seedlings (Boutilier et al., 2002; Chen et al., 2022). Moreover, *AIL5*, *AIL6*, and *AIL7* execute an extra function in phyllotaxy stability and lateral root emergence (Prasad et al., 2011; Hofhuis et al., 2013; Pinon et al., 2013).

The roles of *AIL* genes in panicle, root, seed, leaf, flower, and chloroplast development have been reported in other species. *OsAILs* are involved in panicle branching, panicle structure regulation, and crown root initiation in rice (Kitomi et al., 2011; Harrop et al., 2019; Luong et al., 2021). In poplar, *PtAIL1* plays a positive role in adventitious root formation (Rigal et al., 2012). In *Medicago truncatula*, ectopic expression of *AtANT* under the control of a seed-specific promoter generates larger seeds and improves the germination rate (Confalonieri et al., 2014). In maize, *ZmANT1* regulates leaf and vascular tissue development, chloroplast development, and photosynthesis (Liu et al., 2020). In *Nicotiana tabacum*, *Cucurbita moschata*, *Triticum aestivum*, and *Brassica rapa*, putative orthologs of *ANT* positively regulate organ size. For example, *NtANT* increases the size of leaf and corolla by promoting cell division and expansion (Kuluev et al., 2015), ectopic expression of *CmoANT* accelerates the growth of grafted plants and promotes the size of silique and leaf (Miao et al., 2021), ectopic expression of *TaANT* enlarges plant size by promoting cell proliferation (Zhao et al., 2019b), and ectopic expression of the *BrANT* increases stomatal density and organ size (Ding et al., 2018).

Alfalfa (*Medicago sativa*) has the characteristics of high biomass yield, good forage quality, high adaptability to growing conditions, and palatability for ruminants, and has been called the "Queen of Forage" (Radović et al., 2009). Alfalfa cultivars are allogamous, self-incompatibly, autotetraploid plants with a complex genome, which leads to difficulty in genomics research (Zhu et al., 2005). *Medicago truncatula* has been adopted as a model legume species for a range of genetics and genomics studies. The *AIL* genes have been well studied in many species, but the information and functions of *AIL* genes are largely unknown in legume species. In this study, the genome-wide identification and characterization of the *AIL* gene family was performed in *M. truncatula*. Eleven *MtAIL* genes were identified and their phylogenetic relationship, gene structure, and protein motifs were analyzed. Furthermore, the expression patterns of *MtANTs* showed that *MtANT1* was highly expressed in the SAM and leaf primordium. Loss-of-function mutants of *MtANTs* were isolated and the *mtant1 mtant2 mtant3 mtant4* quadruple mutant was generated. The quadruple mutant exhibited obvious defects in leaf size, while, transgenics overexpressing *MtANT1* produced enlarged leaves. Cellular level analysis indicates that *MtANTs* regulate leaf size mainly through cell proliferation. Our study provides detailed information on the *MtAILs* and demonstrates that *MtANT* genes play vital roles in leaf growth in *M. truncatula*.

Materials and methods

Plant material and growth conditions

Medicago truncatula ecotype R108 was used as the wild type in this study. *mtant1-1*, *mtant2-1*, *mtant3-1*, and *mtant4-1* were identified from a tobacco (*Nicotiana tabacum*) *Tnt1* retrotransposon-tagged mutant population of *M. truncatula* (Tadege et al., 2008). The seeds were scarified with sandpaper and treated at 4°C for 7 days. The germinated seeds were planted in a nursery seedling plate for 3 weeks. Then, the seedlings were transferred to soil and grown at 22°C ± 2°C under long-day conditions (16-h light and 8-h dark), with a relative humidity of 70%–80%.

Identification and phylogenetic analysis of AIL genes in *M. truncatula*

To identify the AIL proteins in *M. truncatula*, 8 AILs in *Arabidopsis thaliana* and 10 AILs in *Oryza sativa* were used to

execute BLASTP Search against the sequence database of the *Medicago truncatula* in Phytozome (https://phytozome-next.jgi.doe.gov/). We selected the fullest transcripts for the study, and other splice variants were excluded. To investigate the phylogenetic relationships of AILs in different species, 8 AIL proteins in *Arabidopsis thaliana*, 10 AIL proteins in *Oryza sativa*, 11 AIL proteins in *Pisum sativum*, 9 AIL proteins in *Lotus japonicus*, 18 AIL proteins in *Glycine max* and 11 identified AIL proteins in *Medicago truncatula* were used to construct the phylogenetic tree. The phylogenetic trees were generated with the Neighbor-Joining method and 1000 Bootstrap Replications using MEGA6.06 (Kumar et al., 2008). The amino acid sequences of the AtAILs were obtained from The Arabidopsis Information Resource (TAIR) database (http://www.Arabidopsis.org/). The OsAILs protein sequences were obtained from the Rice Genome Annotation Project (http://rice.plantbiology.msu.edu/) and National Center for Biotechnology Information (https://www.ncbi.nlm.nih.gov/). The GmAILs and LjAILs protein sequences were obtained in Phytozome, and the PsAILs protein sequences were obtained in the Pea Genome Database (https://www.peagdb.com/index/) (Yang et al., 2022).

Gene structure, conserved domains and motif analysis

Exon and intron structures analysis of MtAIL genes were determined by aligning the CDS sequences and their corresponding genomic DNA sequences using the Gene Structure Display Server (GSDS 2.0, http://gsds.gao-lab.org/) (Wang et al., 2019). Conserved motifs in MtAIL proteins were analyzed with the Multiple Em for Motif Elicitation (MEME, https://meme-suite.org/meme/tools/meme) with the following parameters: maximum of motif width, 80; minimum width of motif, 4; maximum motif number, 10 (Wang et al., 2018). The MtAIL proteins were aligned using Clustal X2 (Larkin et al., 2007), and GeneDoc software was used for homology shading (Brocard-Gifford et al., 2004).

RT-PCR and qRT-PCR analysis and statistical analysis

For gene expression pattern analysis, total RNA was extracted from the leaves, vegetative buds, flowers, stems, petioles, pods, and roots. To analyze the relative expression levels of *MtANT1* in the overexpressing plants, RNA was extracted from 30-d-old mature leaves of wild type and transgenic plants. For RT-PCR analysis, RNA was extracted from vegetative buds of wild type and mutant lines. RNA extraction, cDNA synthesis, qRT-PCR, and RT-PCR analyses were performed as described previously (Zhou et al., 2011). The primers used for qRT-PCR and RT-PCR are listed in Supplementary Table S1. *T*-test was used to compare the means of different populations.

In situ hybridization analysis

For RNA *in situ* hybridization, the 583-bp CDS of *MtANT1*, the 546-bp CDS of *MtANT2*, and the 502-bp CDS of *MtANT3* were amplified. The PCR products were cloned into the pGEM-T vector (Promega). The sense and anti-sense probes were made according to the previous report (Zhao et al., 2019a). 30-d-old wild type vegetative buds were used for RNA *in situ* hybridization as previously described (Zhou et al., 2011). The primers used for RNA *in situ* hybridization are listed in Supplementary Table S1

Plasmids and plant transformation

To obtain the *MtANT1* overexpression construction, the full-length CDS of *MtANT1* was obtained by PCR amplification and inserted into the pENTR/D-TOPO vector (Invitrogen), and then recombined with final vector pEarleyGate 100, using the Gateway LR reactions (Invitrogen) (Earley et al., 2006). The primers used are listed in Supplementary Table S1. The *35S:MtANT1* construct was introduced into *Agrobacterium* strain EHA105. For stable transformation, leaves of wild type were used for transformation (Cosson et al., 2006).

SEM analysis

SEM was performed as described previously (Zhang et al., 2022). Briefly, leaves were fixed, dehydrated, critical point dried, and observed for imaging.

Results

Identification and phylogenetic analysis of AILs in *M. truncatula*

Eleven putative MtAIL proteins were identified by BLAST Search in Phytozome. The length of 11 MtAILs proteins ranged from 402 to 688 amino acids. The gene locus, exon number, amino acid length, molecular weight (Mw), and chromosome location are listed in Table 1. Based on the gene locus, these *MtAIL* genes showed uneven distribution on the *M. truncatula* chromosomes. Chromosome 1, 2, 3, and 7 contained one *MtAIL* gene, respectively. Both chromosome 5 and 8 contained two *MtAIL* genes, chromosome 4 contained four *MtAIL* genes, and no *MtAIL* gene

was located on chromosome 6 (Table 1). To further investigate the evolutionary relationship between MtAIL proteins and homologs in other species, a phylogenetic tree was constructed, including 8 AtAILs, 10 OsAILs, 9 LjAILs, 18 GmAILs, 11 PsAILs and 11 MtAILs (Figure 1). Based on the phylogenetic analysis, all the *MtAIL* genes were named according to their closest *Arabidopsis* orthologs. Furthermore, 67 AIL proteins were classified into six clades: ANT, AIL1, AIL2, AIL3/AIL4, AIL5, and AIL6/7. MtAILs were close to PsAILs, and GmAILs were more closely related to LjAILs. Moreover, every GmAIL gene contained more than one copy. OsAILs were separated from others mainly because of the species differences. ANT clade contained four members of MtAILs which were named MtANT1 to MtANT4. Clades AIL1, AIL2, AIL3/AIL4, AIL5, and AIL6/7 contained seven MtAILs which were named MtAIL1 to MtAIL7 (Figure 1). Phylogenetic analysis also showed that MtAIL1 was more closely related to MtANTs, MtAIL2 was close to MtAIL3 and MtAIL4, MtAIL5 was clustered with MtAIL6 and MtAIL7 (Figure 1).

TABLE 1 *AIL* gene family in *M. truncatula* .

Name	Locus	CDS (nt)	Exons	Length (aa)	MW (kDa)	Chromosome location
MtANT2	Medtr4g097520	1977	9	658	73.8	chr4:40188318..40192060 forward
MtANT1	Medtr1g017400	1995	9	664	74.13	chr1:4844539..4848969 reverse
MtANT3	Medtr3g103460	1986	9	661	73.18	chr3:47751101..47755318 forward
MtAIL1	Medtr8g020510	1722	9	573	63.95	chr8:7209111..7212535 forward
MtANT4	Medtr5g015070	1635	9	544	61.52	chr5:5176272..5179958 reverse
MtAIL5	Medtr4g127930	1557	9	518	55.82	chr4:53232819..53237003 reverse
MtAIL2	Medtr7g080460	2067	9	688	76.57	chr7:30617122..30621534 reverse
MtAIL3	Medtr2g098180	1578	9	525	58.7	chr2:41962850..41966348 reverse
MtAIL4	Medtr4g065370	1644	9	547	61.25	chr4:24560916..24564307 reverse
MtAIL6	Medtr5g031880	1545	9	514	56.95	chr5:13680654..13684967 reverse
MtAIL7	Medtr8g068510	1209	9	402	44.79	chr8:28586113..28591359 reverse

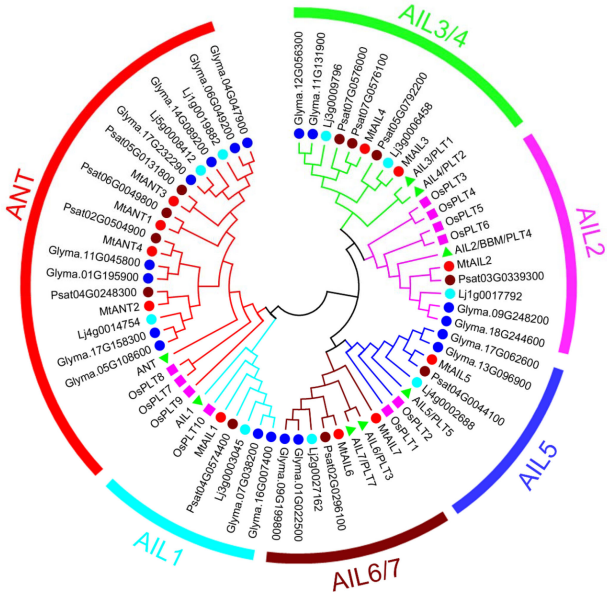


FIGURE 1
Phylogenetic tree analysis of AIL transcription factors from *Medicago truncatula*, *Arabidopsis thaliana*, *Oryza sativa*, *Pisum sativum*, *Lotus japonicus* and *Glycine max*. The Neighbor-Joining tree was constructed using full length protein sequences from *M. truncatula* (11), *Arabidopsis* (8), *Oryza sativa* (10), *Pisum sativum* (11), *Lotus japonicus* (9) and *Glycine max* (18) in MEGA6.06 with 1000 bootstrap replicates. The red, brown, light blue and blue circles indicate *M. truncatula*, *Pisum sativum*, *Lotus japonicus* and *Glycine max*, respectively. The pink boxes and green triangles indicate *Oryza sativa* and *Arabidopsis*.

Gene structures and conserved motifs analysis

To further study the diversification of *MtAILs*, the corresponding gene structures of *MtAILs* were analyzed. The exon/intron organization of *MtAILs* was investigated by aligning the coding sequences and corresponding genomic sequences. All the *MtAILs* displayed 9 exons and 8 introns (Figure 2), suggesting that the gene structures of the *MtAIL* family are conserved during evolution. Previous studies showed that AILs belong to the AP2-like transcription factor subfamily, which is characterized by two AP2 domains (Nole-Wilson et al., 2005; Licausi et al., 2013). According to this, the amino acid sequences of MtAILs were aligned, and two conserved AP2 domains were shown in MtAILs, including the N-terminal AP2-R1 and C-terminal AP2-R2 domains (Supplementary Figure S1). To get a better understanding of the protein sequence characteristics of MtAILs, the motifs of MtAILs were analyzed. Online MEME search was performed and 10 conserved motifs were identified in MtAILs (Figure 3). The motifs 1, 2, 3, 4 were found in each MtAIL. Among them, motifs 1, 2 and 3 were found to be similar to the AP2-R1-linker-AP2-R2 region of AP2-like proteins. Motif 4 was located downstream of the AP2-R2 domain and sequence analysis suggests that it may function as a nuclear localization signal (Aida et al., 2004; Dipp-Álvarez and Cruz-Ramírez, 2019). Motif 5 contained the euANT2 motif (WLGFSLF), and motif 6 was the euANT3 motif (PKLEDFLG). Motif 5 and 6 were conserved motifs of euANT clade protein and existed in almost all the MtAILs (Kim et al., 2006; Dipp-Álvarez and Cruz-Ramírez, 2019). Motif 7 was absent in MtAIL5, 6, 7 proteins, suggesting that motif 7 is lost before the MtAIL5, 6, 7 divergence. Motifs 8 and 9 were specific to MtANT1, 2, 3 and MtAIL1, while motif 10 was only presented in MtANTs, implying that the distribution of motifs among specific groups is related to their functional divergence (Figure 3).

Expression patterns of *MtANTs* in *M. truncatula*

In *Arabidopsis*, *ANT* plays an important role in the regulation of organ initiation, organ development, and cell proliferation (Mizukami and Fischer, 2000; Horstman et al., 2014). In order to investigate the function of *MtANTs*, the expression patterns of *MtANTs* were analyzed. The expression levels of *MtANT1*, *MtANT2*, *MtANT3*, and *MtANT4* were measured by quantitative real-time PCR (qRT-PCR) in different organs, including leaf, vegetative bud, flower, stem, petiole, pod, and root. qRT-PCR results showed that the relative expression levels and patterns of *MtANTs* varied in different organs, but all the *MtANTs* were expressed at the lowest levels in leaves (Figure 4A). Among the four *MtANTs*, the expression of *MtANT1* was much higher than the other three genes in vegetative buds. Additionally, the expression level of *MtANT1* in vegetative buds was higher than other tissues (Figure 4A). To gain spatial information about the expression patterns, *in situ* hybridization was performed for *MtANT1*, *MtANT2*, and *MtANT3*. *In situ* hybridization was not performed for *MtANT4* because it showed very low expression levels in vegetative buds. Strong *MtANT1* signals were detected in SAM, leaf primordia at stage 1 and stage 2, and leaves at stage 7 (Figure 4B). *MtANT2* and *MtANT3* transcripts were not detected in the SAM, and fewer transcripts were detected in leaf primordia and leaves (Figures 4C–D). The overall transcript level of *MtANT1* was much higher than those of *MtANT2* and *MtANT3* (Figures 4B–D). The sense probes were used as the negative controls and did not show any signal (Figures 4E–G).

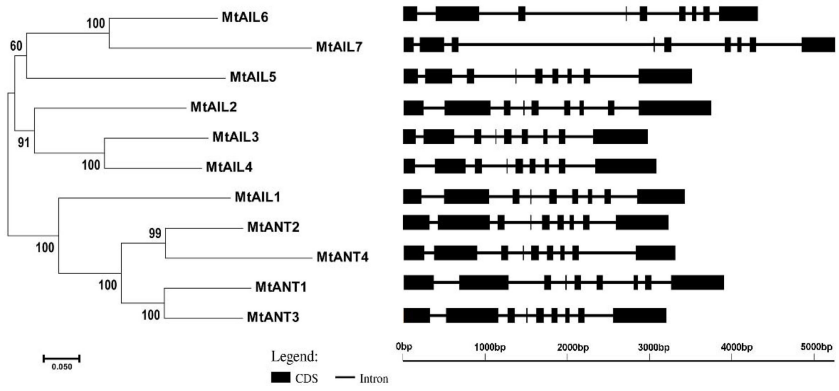


FIGURE 2
The phylogenetic relationship and exon/intron structural analyses of all *AILs* in *M. truncatula*. Introns are represented by black lines and exons are represented by black boxes. Phylogenetic tree was constructed using MEGA6.06 by the Neighbor-Joining method.

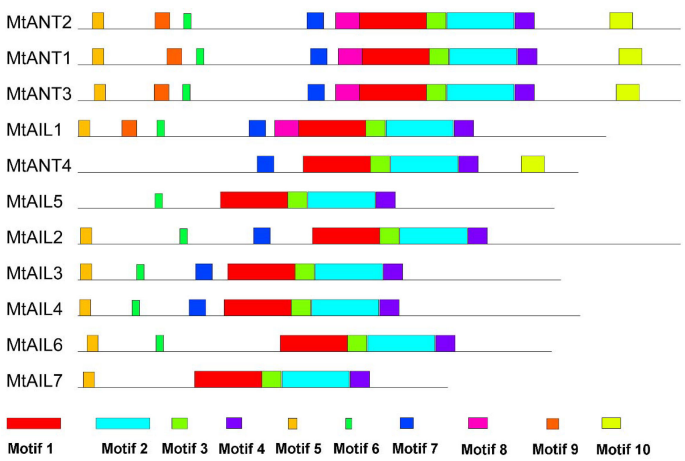


FIGURE 3
Motifs in MtAIL proteins. The motifs in MtAILs were indicated using MEME online tool. Motifs were represented by the boxes with different numbers and colors.

MtANTs are necessary for leaf size maintenance

To investigate the function of *MtANTs* in leaf morphogenesis, a reverse genetic screening was performed on a *Tnt1* retrotransposon-tagged mutant population of *M. truncatula* (Cheng et al., 2014). Insertional mutant alleles were identified in *MtANT1*, *MtANT2*, *MtANT3* and *MtANT4*. Sequence analysis showed that a single *Tnt1* was inserted in the sixth exon of *MtANT1* in *mtant1-1*, the first exon of *MtANT2* in *mtant2-1*, the first exon of *MtANT3* in *mtant3-1*, and the second exon of *MtANT4* in *mtant4-1* (Figures 5A–D). Reverse transcription PCR (RT-PCR) data showed that the transcripts of *MtANT1-4* were interrupted in the *mtant1-1*, *mtant2-1*, *mtant3-1*, and *mtant4-1* mutants, respectiveley (Figures 5E–H). Subsequently, the leaves of mutants were observed. Compared with the wild type, *mtant1-4* mutants did not show obvious defects in leaf morphology and compound leaf pattern (Figures 6A–E). To assess functional redundancy among *MtANTs*, *mtant1-1 mtant3-1* and *mtant2-1 mtant4-1* double mutants were generated. The *mtant1-1 mtant3-1* double mutant exhibited relatively smaller leaves (Figure 6F), but the *mtant2-1 mtant4-1* double mutant did not show obvious differences in leaf

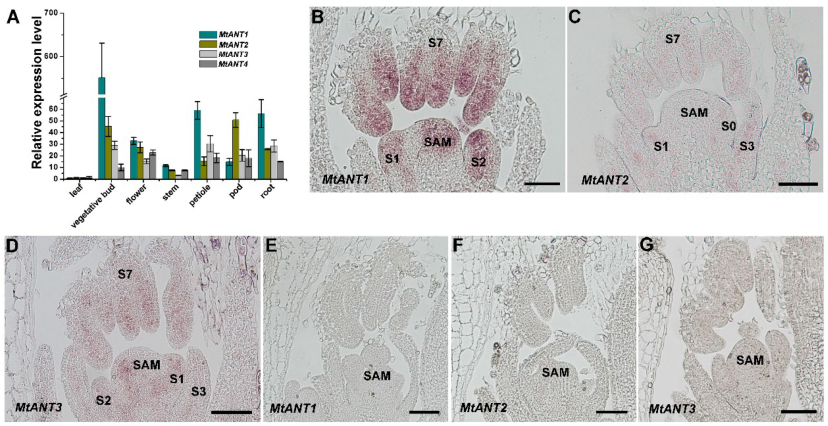


FIGURE 4
Expression patterns of *MtANTs* in *M. truncatula*. **(A)** qRT-PCR analysis of *MtANTs* expression in different organs. *MtUBIQUITIN* was used as the internal control. Bars represent means ± SD of three biological replicates. **(B-D)** *In situ* hybridization analysis of *MtANT1* **(B)**, *MtANT2* **(C)** and *MtANT3* **(D)** in the vegetative buds of wild type. Longitudinal sections of SAM and leaf primordia at the different developmental stages are shown. SAM, shoot apical meristem; S, stage. Bars = 100 μm. **(E–G)** The sense probes were used as the negative controls. SAM, shoot apical meristem; Bars = 100 μm.

pattern compared with wild type (Supplementary Figure S2). Then, high order mutants were generated in the *mtant1-1* background, since *MtANT1* showed the strongest expression in leaf primordia. Simultaneous disruption of *MtANT1*, *2*, *3* or *MtANT1*, *2*, *3*, *4* resulted in a significantly smaller leaf phenotype than that in the wild type (Figures 6G, H). It was worth noting that the leaf phenotypes of *mtant1-1 mtant2-1 mtant3-1* triple mutant and *mtant1-1 mtant2-1 mtant3-1 mtant4-1* quadruple mutant were similar (Figures 6G, H), indicating that *MtANT4* play a limited role in leaf development. Moreover, the length and width of leaves in *mtant1-1 mtant2-1 mtant3-1 mtant4-1* were reduced compared with those of wild type (Figures 6I–L). Compared with the wild type, the length/width ratio was increased, and the leaf area was significantly decreased in *mtant1-1 mtant2-1 mtant3-1 mtant4-1* (Figures 6M–P), demonstrating that simultaneous disruption of *MtANT1*, *2*, *3*, *4* resulted in smaller leaves in *M. truncatula*. To explore the cellular basis for the alteration in leaf dimensions of wild type and *mtant1-1 mtant2-1 mtant3-1 mtant4-1* mutant, we viewed the epidermal cells by scanning electron microscopy (SEM). Results showed that no differences were found for both epidermal cell size and cell number per unit area in the wild type and *mtant* quadruple mutant (Figures 6Q–S), indicating that the smaller leaf area of *mtant* quadruple mutant was resulted from cell proliferation.

MtANT1 is sufficient for increasing leaf size

The aforementioned findings indicated that *MtANTs* play positive roles in regulating leaf size. To determine whether increased expressions of *MtANTs* are sufficient to produce larger leaves, *MtANT1* was chosen to be overexpressed under the control of *CaMV 35S* promoter in the wild type because of its highest expression level in SAM and leaf primordia. Compared with wild type, the expression levels of *MtANT1* were increased by 93- to 669-fold in 35S:*MtANT1* transgenic plants (Supplementary Figure S3). Then, the leaf sizes in *35S:MtANT1* lines and wild type were compared. The *35S:MtANT1-1*, *-6*, and *-8* lines, with the highest expression levels, displayed larger leaves compared with those in wild type (Figures 7A–F). We measured the leaf areas of *35S:MtANT1* transgenic plants and wild type. The transgenic lines showed an increase in the area of leaves in comparison with those in wild type plants (Figures 7G–I). In addition, the length, width and length/width ratio of leaves of *35S:MtANT1-6* transgenic line were measured. The results showed that the length of leaves in *35S:MtANT1-6* was similar to that in the wild type, the leaves of transgenic plant were wider than wild type, and the length/width

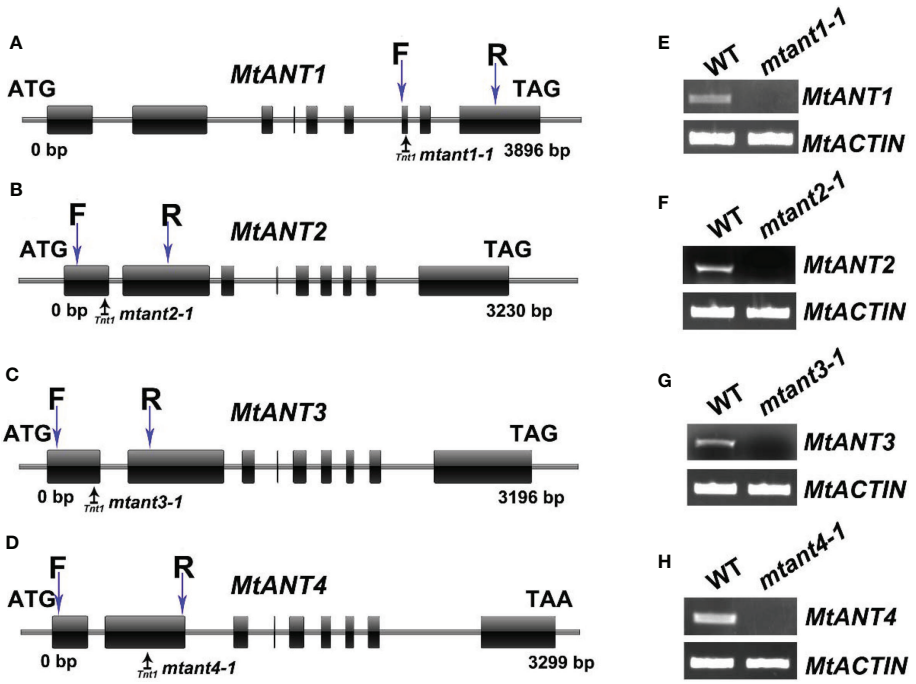


FIGURE 5
Mutant screening of *MtANTs* in *M. truncatula*. **(A–D)** Schematic diagram of the gene structures of *MtANT1-MtANT4*. The positions of the ATG start and TGA/TAA stop codons are shown. Black vertical arrows mark the location of *Tnt1* retrotransposons. Blue vertical arrows mark the location of primers used for RT-PCR. Exon is represented by a box, and intron is represented by a line. **(E–H)** RT-PCR shows the transcripts of *MtANTs* in wild type and *mtant* mutants. *MtACTIN* was used as the control.

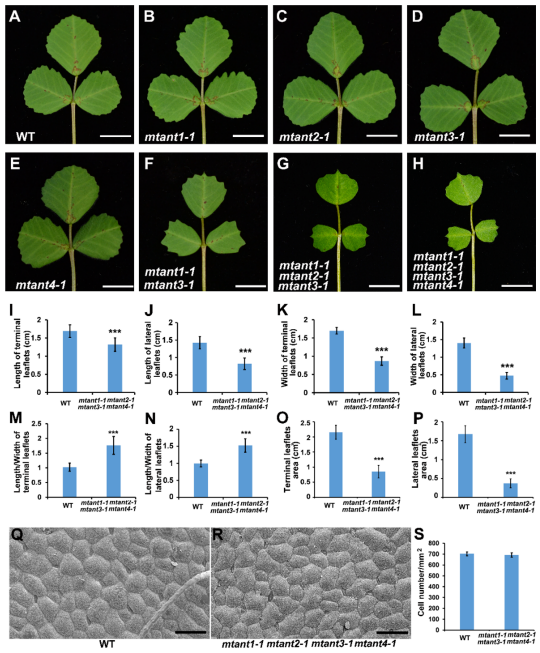


FIGURE 6
Phenotype analyses of leaves in mtant mutants. **(A–H)** Leaf phenotypes of wild type **(A)**, mtant1-1 **(B)**, mtant2-1 **(C)**, mtant3-1 **(D)**, mtant4-1 **(E)**, mtant1-1 mtant3-1 **(F)**, mtant1-1 mtant2-1 mtant3-1 **(G)** and mtant1-1 mtant2-1 mtant3-1 mtant4-1 **(H)**. Bars = 1 cm. **(I–L)** Leaf length **(I–J)** and width **(K, L)** in terminal leaflets and lateral leaflets of wild type and mtant1-1 mtant2-1 mtant3-1 mtant4-1. **(M–P)** Length/width ratio **(M, N)** and leaf area **(O, P)** of terminal leaflets and lateral leaflets in wild type and mtant1-1 mtant2-1 mtant3-1 mtant4-1. **(Q, R)** Scanning electron microscopy analyses of the adaxial surface of leaves in wild type **(Q)** and mtant1-1 mtant2-1 mtant3-1 mtant4-1 mutant **(R)**. Bars = 50 μm. **(S)** Cell number per unit area of the adaxial surface of leaves in the wild type and mtant1-1 mtant2-1 mtant3-1 mtant4-1 mutant. Bars represent means ± SD. ***P < 0.001.

ratio of *35S:MtANT1-6* was decreased (Figures 7J–L). These observations suggest that ectopic expression of *MtANT1* is able to promote the leaf width and leaf size in *M. truncatula*. The epidermal cells of wild type and *35S:MtANT1-6* plants were also analyzed by SEM. The data showed that the cells size and cell number per unit area were similar between wild type and *35S:MtANT1-6* plants (Figures 7M–O), further demonstrating that the larger leaf area of *35S:MtANT1* transgenic plants was resulted from increased cell proliferation. Overall, *MtANTs* control leaf growth by promoting cell proliferation rather than cell expansion.

Discussion

AINTEGUMENTA-LIKE (AIL) proteins belong to the AP2-like family, and play vital roles in plant developmental process and stress response (Horstman et al., 2014; Meng et al., 2015a; Meng et al., 2015b). AIL transcription factors have been extensively studied in different species, but their functions in the model legume, *M. truncatula*, are largely unknown. According to the phylogenetic analysis, four MtANT genes are identified. MtANT1 and MtANT3 are clustered in one clade, while MtANT2 and MtANT4 are clustered in another clade, indicating the tandem duplication followed by genomic reshuffling in *M. truncatula*. The MtAIL proteins are more closely related to *Pisum sativum* homologs, indicating that their *MtAIL* and *PsAIL* genes may diverge from a common ancestor. Compared with other AILs, most *GmAIL* genes are presented in multiple copies, suggesting that these genes are a product of whole-genome duplication events and relatively slow process of diploidization during the evolutionary process. *AIL2/BBM/PLT4*, *AIL3/PLT1*, *AIL4/PLT2* and *AIL6/PLT3* were reported to regulate the root stem cell niche patterning in *Arabidopsis*, and *OsPLT1-6* are specifically expressed in the primodium of crown root and lateral root in rice (Galinha et al., 2007; Li and Xue, 2011). Based on the phylogenetic tree analysis, we speculate that MtAIL2, 3 and 4 and their homologous proteins in the same clade among leguminous species may be involved in the regulation of root and root nodule development. Gene structure is an important indicator for gene function and classification, of which intron gain or loss is the consequence of selection pressures during evolution (Mattick, 1994). In our study, all the *MtAIL* genes evolve into the same exon-intron structures, further supporting their close evolutionary relationship. Motifs and domains are involved in various regulatory processes including interactions between proteins, transcriptional activity, and DNA binding (Liu et al., 1999). The numbers and distribution of motifs in MtAILs are different, implying that MtAIL members have some differences in function. However, two AP2 domains are highly conserved among the MtAIL proteins, suggesting that the AP2 domains are evolutionarily conserved and necessary for the correct structure of AIL proteins. Moreover, proteins sharing the unique motifs in one cluster are likely to exert similar functions (Du et al., 2012; Zhao et al., 2019b). A unique motif (motif 10) was displayed in four MtANT members. This finding raises a question on whether motif 10 is related to the leaf growth, and future characterization of the function of motif 10 will clarify this point.

AtANT participates in organ size control, floral organ initiation and development (Mizukami and Fischer, 2000; Horstman et al., 2014; Manchado-Rojo et al., 2014), the *ANT*-homologous genes are associated with panicle branching, panicle structure and inflorescence development in rice (Kitomi et al., 2011; Luong et al., 2021). But we observed that *MtANTs* only affect the leaf growth in *M. truncatula*. The *mtant* quadruple mutant and *MtANT1* overexpression plants didn't exhibit other phenotypic changes, such as plant height, floral organ size and inflorescence structure. In addition, MtAIL1 is closer to MtANTs, and the developmental defects of *mtant* quadruple mutant may be masked by MtAIL1. Thus, the multiple

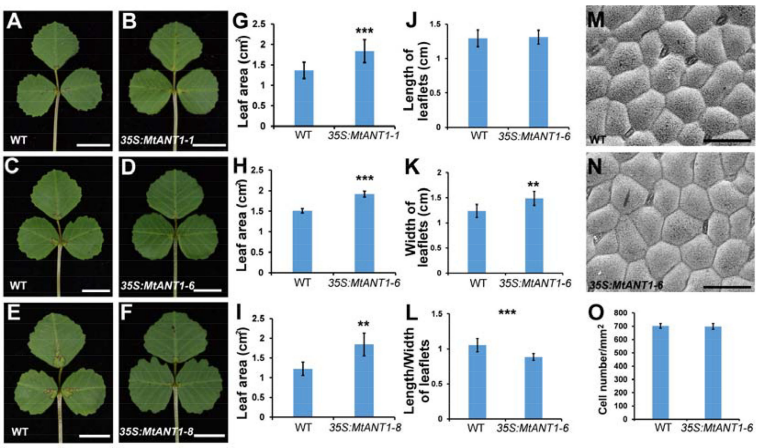


FIGURE 7
Phenotype analyses of leaves in *35S:MtANT1* plants. **(A–F)** Leaf phenotypes of wild type **(A, C, E)** and *35S:MtANT1-1, -6, -8* transgenic plants **(B, D, F)**. Bars = 1 cm. **(G–I)** Leaf area of wild type and *35S:MtANT1-1, -6, -8* plants. **(J–L)** Length **(J)**, width **(K)** and length/width ratio **(L)** of leaves in wild type and *35S:MtANT1-6* plants. **(M–N)** Scanning electron microscopy analyses of the adaxial surface of leaves in wild type **(M)** and *35S:MtANT1-6* plants **(N)** Bars = 50 μm. **(O)** Cell number per unit area of the adaxial surface of leaves in the wild type and *35S:MtANT1-6* plants. Bars represent means ± SD. ***P < 0.001, **P < 0.01.

mutations of *MtAIL1* and *MtANTs* may lead to severe developmental defects. In *Arabidopsis*, *N. tabacum*, *C. moschata*, *T. aestivum*, and *B. rapa*, ectopic expression of *ANT* enlarged the size of leaves (Mizukami and Fischer, 2000; Ding et al., 2018; Zhao et al., 2019b; Miao et al., 2021). In accordance with these reports, a similar phenotype of larger leaves was shown in *35S:MtANT1* plants. These results indicate that *ANT* genes also exert conserved function in increasing leaf area among different species. Biomass is critical for the evaluation of forage grass quality. Therefore, overexpression of *ANT* in legume forages, such as alfalfa, will be helpful to improve forage production.

Previous study showed that *ANGUSTIFOLIA3* (*AN3*) functions as a transcriptional activator of the GRF-INTERACTING FACTOR (GIF) family, and it is probably a potential target of *ANT* in promoting organ growth in *Arabidopsis* (Krizek et al., 2020). The loss-of-function mutant of *AN3* exhibits smaller leaf size, while ectopic expression of *AN3* results in larger leaves (Kim and Kende, 2004; Horiguchi et al., 2005; Lee et al., 2009). The potential regulatory relationship and phenotypes in *Arabidopsis* remind us that the paralogous gene of *AN3* in *M. truncatula* may play a similar role. In addition, *CYCD3* genes encode D-type cell cycle proteins which play key roles in the switch from cell proliferation to cell differentiation (Dewitte et al., 2003; Menges et al., 2006). The expression of *CYCD3;1* is prolonged in the leaves of *35S:ANT* plant to maintain the meristematic competence of cells during organogenesis in *Arabidopsis* (Mizukami and Fischer, 2000). Similarly, *CYCD3.2* is also the downstream target of the AIL1 transcription factor in poplar (Karlberg et al., 2011). According to these reports, it raises the possibility that *MtANTs* regulate the cell cycle genes to determine the leaf size.

The leaf development process includes three intertwined stages: leaf initiation in the SAM, primary morphogenesis, and secondary morphogenesis in which expansion and proliferation of cells occur (Poethig, 1997; Dengler and Tsukaya, 2001; Shani et al., 2009; Bar and Ori, 2014). Leaf complexity is determined during primary morphogenesis. *M. truncatula* is a compound-leafed species, whose adult leaves are trifoliate. In this study, we found the leaf complexities in both quadruple mutant and *MtANT1*-overexpressing plants were unchanged. Therefore, we propose that *MtANTs* mainly regulate the secondary morphogenesis of leaves in *M. truncatula*.

In addition, auxin plays an important role in leaf development by influencing cell proliferation and cell expansion (Tadege et al., 2011; Wang et al., 2021). In *Arabidopsis*, overexpression of the auxin-inducible gene *AUXIN-REGULATED GENE INVOLVED IN ORGAN SIZE* (*ARGOS*) generates larger leaves and prolongs the expression of *ANT*. *ARGOS* functions downstream of the auxin signaling pathway and upstream of *ANT* in the regulation of leaf size (Hu et al., 2003). So, the relationship among auxin related pathways, *MtANTs* and *MtARGOS* should be investigated in the future.

Conclusion

In this study, we performed genome-wide analyses and identified *AIL* genes in *M. truncatula*. We characterized *MtANT* genes expression profiles in different tissues,

suggesting that *MtANT* genes play important roles in the leaf morphogenesis of *M. truncatula*. Simultaneous disruption of *MtANTs* resulted in smaller leaves and overexpression of *MtANT1* led to larger leaves, demonstrating that *MtANTs* are vital for leaf size maintenance. However, they cann't influence the leaf complexity. Further study is needed to elucidate the molecular mechanism of *MtANTs* that are involved in leaf size development.

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

XW, CZ and LH designed the research. XW, JJZ and JZ performed the research and analyzed the data. XW and LH wrote the paper. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by grants from the National Natural Science Foundation of China (U1906201 and 31900172) and Shandong Province (ZR2020KC018 and ZR2020QC035).

Acknowledgments

We thank Prof. Kirankumar Mysore (Oklahoma state university), Dr. Jiangqi Wen (Oklahoma state university) for providing the *Tnt1* mutants, and Haiyan Yu from State Key Laboratory of Microbial Technology of Shandong University for help and guidance in microscope.

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.1050462/full#supplementary-material

References

Aida, M., Beis, D., Heidstra, R., Willemsen, V., Blilou, I., Galinha, C., et al. (2004). The PLETHORA genes mediate patterning of the arabidopsis root stem cell niche. *Cell* 119 (1), 109–120. doi: 10.1016/j.cell.2004.09.018

Bar, M., and Ori, N. (2014). Leaf development and morphogenesis. *Development* 141 (22), 4219–4230. doi: 10.1242/dev.106195

Boutilier, K., Offringa, R., Sharma, V. K., Kieft, H., Ouellet, T., Zhang, L., et al. (2002). Ectopic expression of BABY BOOM triggers a conversion from vegetative to embryonic growth. *Plant Cell* 14 (8), 1737–1749. doi: 10.1105/tpc.001941

Brocard-Gifford, I., Lynch, T. J., Garcia, M. E., Malhotra, B., and Finkelstein, R. R. (2004). The arabidopsis thaliana ABSCISIC ACID-INSENSITIVE8 encodes a novel protein mediating abscisic acid and sugar responses essential for growth. *Plant Cell* 16 (2), 406–421. doi: 10.1105/tpc.018077

Bui, L. T., Pandzic, D., Youngstrom, C. E., Wallace, S., Irish, E. E., Szövényi, P., et al. (2017). A fern AINTEGUMENTA gene mirrors BABY BOOM in promoting apogamy in ceratopteris richardii. *Plant J.* 90 (1), 122–132. doi: 10.1111/tpj.13479

Cheng, X., Wang, M., Lee, H. K., Tadege, M., Ratet, P., Udvardi, M., et al. (2014). An efficient reverse genetics platform in the model legume medicago truncatula. *New Phytol.* 201 (3), 1065–1076. doi: 10.1111/nph.12575

Chen, B., Maas, L., Figueiredo, D., Zhong, Y., Reis, R., Li, M., et al. (2022). BABY BOOM regulates early embryo and endosperm development. *Proc. Natl. Acad. Sci. U.S.A.* 119 (25), e2201761119. doi: 10.1073/pnas.2201761119

Confalonieri, M., Carelli, M., Galimberti, V., Macovei, A., Panara, F., Biggiogera, M., et al. (2014). Seed-specific expression of AINTEGUMENTA in medicago truncatula led to the production of larger seeds and improved seed germination. *Plant Mol. Biol. Rep.* 32 (5), 957–970. doi: 10.1007/s11105-014-0706-4

Cosson, V., Durand, P., d'Erfurth, I., Kondorosi, A., and Ratet, P. (2006). Medicago truncatula transformation using leaf explants. *Methods Mol. Biol.* 343, 115–127. doi: 10.1385/1-59745-130-4:115

Dengler, N. G., and Tsukaya, H. (2001). Leaf morphogenesis in dicotyledons: current issues. *Int. J. Plant Sci.* 162 (3), 459–464. doi: 10.1086/320145

Dewitte, W., Riou-Khamlichi, C., Scofield, S., Healy, J. M., Jacqmard, A., Kilby, N. J., et al. (2003). Altered cell cycle distribution, hyperplasia, and inhibited differentiation in arabidopsis caused by the d-type cyclin CYCD3. *Plant Cell* 15 (1), 79–92. doi: 10.1105/tpc.004838

Ding, Q., Cui, B., Li, J., Li, H., Zhang, Y., Lv, X., et al. (2018). Ectopic expression of a brassica rapa AINTEGUMENTA gene (BrANT-1) increases organ size and stomatal density in arabidopsis. *Sci. Rep.* 8 (1), 10528. doi: 10.1038/s41598-018-28606-4

Dipp-Álvarez, M., and Cruz-Ramírez, A. (2019). A phylogenetic study of the ANT family points to a preANT gene as the ancestor of basal and euANT transcription factors in land plants. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.00017

Du, H., Yang, S. S., Liang, Z., Feng, B. R., Liu, L., Huang, Y. B., et al. (2012). Genome-wide analysis of the MYB transcription factor superfamily in soybean. *BMC Plant Biol.* 12, 106. doi: 10.1186/1471-2229-12-106

Earley, K. W., Haag, J. R., Pontes, O., Opper, K., Juehne, T., Song, K., et al. (2006). Gateway-compatible vectors for plant functional genomics and proteomics. *Plant J.* 45 (4), 616–629. doi: 10.1111/j.1365-313X.2005.02617.x

Galinha, C., Hofhuis, H., Luijten, M., Willemsen, V., Blilou, I., Heidstra, R., et al. (2007). PLETHORA proteins as dose-dependent master regulators of arabidopsis root development. *Nature* 449 (7165), 1053–1057. doi: 10.1038/nature06206

Han, X., Liu, K., Yuan, G., He, S., Cong, P., and Zhang, C. (2022). Genome-wide identification and characterization of AINTEGUMENTA-LIKE (AIL) family genes in apple (Malus domestica borkh.). *Genomics* 114 (2), 110313. doi: 10.1016/j.ygeno.2022.110313

Harrop, T. W. R., Mantegazza, O., Luong, A. M., Béthune, K., Lorieux, M., Jouannic, S., et al. (2019). A set of AP2-like genes is associated with inflorescence branching and architecture in domesticated rice. *J. Exp. Bot.* 70 (20), 5617–5629. doi: 10.1093/jxb/erz340

Hofhuis, H., Laskowski, M., Du, Y., Prasad, K., Grigg, S., Pinon, V., et al. (2013). Phyllotaxis and rhizotaxis in arabidopsis are modified by three PLETHORA transcription factors. *Curr. Biol.* 23 (11), 956–962. doi: 10.1016/j.cub.2013.04.048

Horiguchi, G., Kim, G. T., and Tsukaya, H. (2005). The transcription factor AtGRF5 and the transcription coactivator AN3 regulate cell proliferation in leaf primordia of arabidopsis thaliana. *Plant J.* 43 (1), 68–78. doi: 10.1111/j.1365-313X.2005.02429.x

Horstman, A., Willemsen, V., Boutilier, K., and Heidstra, R. (2014). AINTEGUMENTA-LIKE proteins: hubs in a plethora of networks. *Trends Plant Sci.* 19 (3), 146–157. doi: 10.1016/j.tplants.2013.10.010

Hu, Y., Xie, Q., and Chua, N. H. (2003). The arabidopsis auxin-inducible gene ARGOS controls lateral organ size. *Plant Cell* 15 (9), 1951–1961. doi: 10.1105/tpc.013557

Karlberg, A., Bako, L., and Bhalerao, R. P. (2011). Short day-mediated cessation of growth requires the downregulation of AINTEGUMENTALIKE1 transcription factor in hybrid aspen. *PloS Genet.* 7 (11), e1002361. doi: 10.1371/journal.pgen.1002361

Kim, J. H., and Kende, H. (2004). A transcriptional coactivator, AtGIF1, is involved in regulating leaf growth and morphology in arabidopsis. *Proc. Natl. Acad. Sci. U.S.A.* 101 (36), 13374–13379. doi: 10.1073/pnas.0405450101

Kim, S., Soltis, P. S., Wall, K., and Soltis, D. E. (2006). Phylogeny and domain evolution in the APETALA2-like gene family. *Mol. Biol. Evol.* 23 (1), 107–120. doi: 10.1093/molbev/msj014

Kitomi, Y., Ito, H., Hobo, T., Aya, K., Kitano, H., and Inukai, Y. (2011). The auxin responsive AP2/ERF transcription factor CROWN ROOTLESS5 is involved in crown root initiation in rice through the induction of OsRR1, a type-a response regulator of cytokinin signaling. *Plant J.* 67 (3), 472–484. doi: 10.1111/j.1365-313X.2011.04610.x

Klucher, K. M., Chow, H., Reiser, L., and Fischer, R. L. (1996). The AINTEGUMENTA gene of arabidopsis required for ovule and female gametophyte development is related to the floral homeotic gene APETALA2. *Plant Cell* 8 (2), 137–153. doi: 10.1105/tpc.8.2.137

Krizek, B. (2009). AINTEGUMENTA and AINTEGUMENTA-LIKE6 act redundantly to regulate arabidopsis floral growth and patterning. *Plant Physiol.* 150 (4), 1916–1929. doi: 10.1104/pp.109.141119

Krizek, B. A. (2015). AINTEGUMENTA-LIKE genes have partly overlapping functions with AINTEGUMENTA but make distinct contributions to arabidopsis thaliana flower development. *J. Exp. Bot.* 66 (15), 4537–4549. doi: 10.1093/jxb/erv224

Krizek, B. A., Bantle, A. T., Heflin, J. M., Han, H., Freese, N. H., and Loraine, A. E. (2021). AINTEGUMENTA and AINTEGUMENTA-LIKE6 directly regulate floral homeotic, growth, and vascular development genes in young arabidopsis flowers. *J. Exp. Bot.* 72 (15), 5478–5493. doi: 10.1093/jxb/erab223

Krizek, B. A., Blakley, I. C., Ho, Y. Y., Freese, N., and Loraine, A. E. (2020). The arabidopsis transcription factor AINTEGUMENTA orchestrates patterning genes and auxin signaling in the establishment of floral growth and form. *Plant J.* 103 (2), 752–768. doi: 10.1111/tpj.14769

Kuluev, B., Avalbaev, A., Nurgaleeva, E., Knyazev, A., Nikonorov, Y., and Chemeris, A. (2015). Role of AINTEGUMENTA-like gene NtANTL in the regulation of tobacco organ growth. *J. Plant Physiol.* 189, 11–23. doi: 10.1016/j.jplph.2015.08.009

Kumar, S., Nei, M., Dudley, J., and Tamura, K. (2008). MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief Bioinform.* 9 (4), 299–306. doi: 10.1093/bib/bbn017

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., et al. (2007). Clustal W and clustal X version 2.0. *Bioinformatics* 23 (21), 2947–2948. doi: 10.1093/bioinformatics/btm404

Lee, B. H., Ko, J. H., Lee, S., Lee, Y., Pak, J. H., and Kim, J. H. (2009). The arabidopsis GRF-INTERACTING FACTOR gene family performs an overlapping function in determining organ size as well as multiple developmental properties. *Plant Physiol.* 151 (2), 655–668. doi: 10.1104/pp.109.141838

Licausi, F., Ohme-Takagi, M., and Perata, P. (2013). APETALA2/Ethylene responsive factor (AP2/ERF) transcription factors: mediators of stress responses and developmental programs. *New Phytol.* 199 (3), 639–649. doi: 10.1111/nph.12291

Liu, W. Y., Lin, H. H., Yu, C. P., Chang, C. K., Chen, H. J., Lin, J. J., et al. (2020). Maize ANT1 modulates vascular development, chloroplast development, photosynthesis, and plant growth. *Proc. Natl. Acad. Sci. U.S.A.* 117 (35), 21747–21756. doi: 10.1073/pnas.2012245117

Liu, L., White, M. J., and MacRae, T. H. (1999). Transcription factors and their genes in higher plants functional domains, evolution and regulation. *Eur. J. Biochem.* 262 (2), 247–257. doi: 10.1046/j.1432-1327.1999.00349.x

Li, P., and Xue, H. (2011). Structural characterization and expression pattern analysis of the rice PLT gene family. *Acta Biochim. Biophys. Sin. (Shanghai)* 43 (9), 688–697. doi: 10.1093/abbs/gmr068

Luong, A. M., Adam, H., Gauron, C., Affortit, P., Ntakirutimana, F., Khong, N. G., et al. (2021). Functional diversification of euANT/PLT genes in oryza sativa panicle architecture determination. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.692955

Manchado-Rojo, M., Weiss, J., and Egea-Cortines, M. (2014). Validation of aintegumenta as a gene to modify floral size in ornamental plants. *Plant Biotechnol. J.* 12 (8), 1053–1065. doi: 10.1111/pbi.12212

Mattick, J. S. (1994). Introns: evolution and function. *Curr. Opin. Genet. Dev.* 4 (6), 823–831. doi: 10.1016/0959-437x(94)90066-3

Menges, M., Samland, A. K., Planchais, S., and Murray, J. A. (2006). The d-type cyclin CYCD3;1 is limiting for the G1-to-S-phase transition in arabidopsis. *Plant Cell* 18 (4), 893–906. doi: 10.1105/tpc.105.039636

Meng, L. S., Wang, Y. B., Yao, S. Q., and Liu, A. (2015a). Arabidopsis AINTEGUMENTA mediates salt tolerance by trans-repressing SCABP8. *J. Cell Sci.* 128 (15), 2919–2927. doi: 10.1242/jcs.172072

Meng, L. S., Wang, Z. B., Yao, S. Q., and Liu, A. (2015b). The ARF2-ANT-COR15A gene cascade regulates ABA-signaling-mediated resistance of large seeds to drought in arabidopsis. *J. Cell Sci.* 128 (21), 3922–3932. doi: 10.1242/jcs.171207

Miao, L., Li, S. Z., Shi, A. K., Li, Y. S., He, C. X., Yan, Y., et al. (2021). Genome-wide analysis of the AINTEGUMENTA-like (AIL) transcription factor gene family in pumpkin (Cucurbita moschata duch.) and CmoANT1.2 response in graft union healing. *Plant Physiol. Biochem.* 162, 706–715. doi: 10.1016/j.plaphy.2021.03.036

Mizukami, Y., and Fischer, R. L. (2000). Plant organ size control: AINTEGUMENTA regulates growth and cell numbers during organogenesis. *Proc. Natl. Acad. Sci. U.S.A.* 97 (2), 942–947. doi: 10.1073/pnas.97.2.942

Mizumoto, K., Hatano, H., Hirabayashi, C., Murai, K., and Takumi, S. (2009). Altered expression of wheat AINTEGUMENTA homolog, WANT-1, in pistil and pistil-like transformed stamen of an alloplasmic line with aegilops crassa cytoplasm. *Dev. Genes Evol.* 219 (4), 175–187. doi: 10.1007/s00427-009-0275-y

Mudunkothge, J. S., and Krizek, B. A. (2012). Three arabidopsis AIL/PLT genes act in combination to regulate shoot apical meristem function. *Plant J.* 71 (1), 108–121. doi: 10.1111/j.1365-313X.2012.04975.x

Nole-Wilson, S., and Krizek, B. A. (2006). AINTEGUMENTA contributes to organ polarity and regulates growth of lateral organs in combination with YABBY genes. *Plant Physiol.* 141 (3), 977–987. doi: 10.1104/pp.106.076604

Nole-Wilson, S., Tranby, T. L., and Krizek, B. A. (2005). AINTEGUMENTA-like (AIL) genes are expressed in young tissues and may specify meristematic or division-competent states. *Plant Mol. Biol.* 57 (5), 613–628. doi: 10.1007/s11103-005-0955-6

Pinon, V., Prasad, K., Grigg, S. P., Sanchez-Perez, G. F., and Scheres, B. (2013). Local auxin biosynthesis regulation by PLETHORA transcription factors controls phyllotaxis in arabidopsis. *Proc. Natl. Acad. Sci. U.S.A.* 110 (3), 1107–1112. doi: 10.1073/pnas.1213497110

Poethig, R. S. (1997). Leaf morphogenesis in flowering plants. *Plant Cell* 9 (7), 1077–1087. doi: 10.1105/tpc.9.7.1077

Prasad, K., Grigg, S. P., Barkoulas, M., Yadav, R. K., Sanchez-Perez, G. F., Pinon, V., et al. (2011). Arabidopsis PLETHORA transcription factors control phyllotaxis. *Curr. Biol.* 21 (13), 1123–1128. doi: 10.1016/j.cub.2011.05.009

Radović, J., Sokolović, D., and Marković, J. (2009). Alfalfa-most important perennial forage legume in animal husbandry. *Biotechnol. Anim. Husbandry* 25 (5-6-1), 465–475. doi: 10.2298/BAH0906465R

Rigal, A., Yordanov, Y. S., Perrone, I., Karlberg, A., Tisserant, E., Bellini, C., et al. (2012). The AINTEGUMENTA LIKE1 homeotic transcription factor PtAIL1 controls the formation of adventitious root primordia in poplar. *Plant Physiol.* 160 (4), 1996–2006. doi: 10.1104/pp.112.204453

Shani, E., Burko, Y., Ben-Yaakov, L., Berger, Y., Amsellem, Z., Goldshmidt, A., et al. (2009). Stage-specific regulation of solanum lycopersicum leaf maturation by class 1 KNOTTED1-LIKE HOMEOBOX proteins. *Plant Cell* 21 (10), 3078–3092. doi: 10.1105/tpc.109.068148

Shen, S., Sun, F., Zhu, M., Chen, S., Guan, M., Chen, R., et al. (2020). Genome-wide identification AINTEGUMENTA-like (AIL) genes in brassica species and

expression patterns during reproductive development in. *PloS One* 15 (6), e0234411. doi: 10.1371/journal.pone.0234411

Tadege, M., Lin, H., Bedair, M., Berbel, A., Wen, J., Rojas, C. M., et al. (2011). STENOFOLIA regulates blade outgrowth and leaf vascular patterning in medicago truncatula and nicotiana sylvestris. *Plant Cell* 23 (6), 2125–2142. doi: 10.1105/tpc.111.085340

Tadege, M., Wen, J., He, J., Tu, H., Kwak, Y., Eschstruth, A., et al. (2008). Large-Scale insertional mutagenesis using the Tnt1 retrotransposon in the model legume medicago truncatula. *Plant J.* 54 (2), 335–347. doi: 10.1111/j.1365-313X.2008.03418.x

Wang, H., Kong, F., and Zhou, C. (2021). From genes to networks: The genetic control of leaf development. *J. Integr. Plant Biol.* 63 (7), 1181–1196. doi: 10.1111/jipb.13084

Wang, H., Lu, Z., Xu, Y., Kong, L., Shi, J., Liu, Y., et al. (2019). Genome-wide characterization of SPL family in medicago truncatula reveals the novel roles of miR156/SPL module in spiky pod development. *BMC Genomics* 20 (1), 552. doi: 10.1186/s12864-019-5937-1

Wang, H., Wang, H., Liu, R., Xu, Y., Lu, Z., and Zhou, C. (2018). Genome-wide identification of TCP family transcription factors in medicago truncatula reveals significant roles of miR319-targeted TCPs in nodule development. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.00774

Yamaguchi, N., Jeong, C. W., Nole-Wilson, S., Krizek, B. A., and Wagner, D. (2016). AINTEGUMENTA and AINTEGUMENTA-LIKE6/PLETHORA3 induce LEAFY expression in response to auxin to promote the onset of flower formation in arabidopsis. *Plant Physiol.* 170 (1), 283–293. doi: 10.1104/pp.15.00969

Yang, T., Liu, R., Luo, Y., Hu, S., Wang, D., Wang, C., et al. (2022). Improved pea reference genome and pan-genome highlight genomic features and evolutionary characteristics. *Nat. Genet.* 54 (10), 1553–1563. doi: 10.1038/s41588-022-01172-2

Zhang, J., Wang, X., Han, L., Zhang, J., Xie, Y., Li, J., et al. (2022). The formation of stipule requires the coordinated actions of the legume orthologs of arabidopsis BLADE-ON-PETIOLE and LEAFY. *New Phytol.* 236(4), 1512–1528. doi: 10.1111/nph.18445

Zhao, Y., Liu, R., Xu, Y., Wang, M., Zhang, J., Bai, M., et al. (2019a). AGLF provides c-function in floral organ identity through transcriptional regulation of AGAMOUS in medicago truncatula. *Proc. Natl. Acad. Sci. U.S.A.* 116 (11), 5176–5181. doi: 10.1073/pnas.1820468116

Zhao, Y., Ma, R., Xu, D., Bi, H., Xia, Z., and Peng, H. (2019bb). Genome-wide identification and analysis of the AP2 transcription factor gene family in wheat (Triticum aestivum l.). *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.01286

Zhou, C., Han, L., Hou, C., Metelli, A., Qi, L., Tadege, M., et al. (2011). Developmental analysis of a medicago truncatula smooth leaf margin1 mutant reveals context-dependent effects on compound leaf development. *Plant Cell* 23 (6), 2106–2124. doi: 10.1105/tpc.111.085464

Zhu, H., Choi, H. K., Cook, D. R., and Shoemaker, R. C. (2005). Bridging model and crop legumes through comparative genomics. *Plant Physiol.* 137 (4), 1189–1196. doi: 10.1104/pp.104.058891

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 11 November 2022
DOI 10.3389/fpls.2022.1052017

OPEN ACCESS

EDITED BY
Zhengjun Xia,
Chinese Academy of Sciences (CAS), China

REVIEWED BY
Dawei Xin,
Northeast Agricultural University, China
Pascal Ratet,
UMR9213 Institut des Sciences des Plantes de Paris Saclay (IPS2), France

*CORRESPONDENCE
Liyu Chen
chenliyu1715@gzhu.edu.cn
Fanjiang Kong
kongfj@gzhu.edu.cn

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Functional and Applied Plant Genomics, a section of the journal Frontiers in Plant Science

RECEIVED 23 September 2022
ACCEPTED 25 October 2022
PUBLISHED 11 November 2022

CITATION
Zhang Y, Cheng Q, Liao C, Li L, Gou C, Chen Z, Wang Y, Liu B, Kong F and Chen L (2022) GmTOC1b inhibits nodulation by repressing *GmNIN2a* and *GmENOD40-1* in soybean. *Front. Plant Sci.* 13:1052017. doi: 10.3389/fpls.2022.1052017

GmTOC1b inhibits nodulation by repressing *GmNIN2a* and *GmENOD40-1* in soybean

Yuhang Zhang†, Qun Cheng†, Chunmei Liao, Lanxin Li, Chuanjie Gou, Zheng Chen, Yanan Wang, Baohui Liu, Fanjiang Kong* and Liyu Chen*

Guangdong Key Laboratory of Plant Adaptation and Molecular Design, Guangzhou Key Laboratory of Crop Gene Editing, Innovative Center of Molecular Genetics and Evolution, School of Life Sciences, Guangzhou University, Guangzhou Higher Education Mega Center, Guangzhou, China

Symbiotic nitrogen fixation is an important factor affecting the yield and quality of leguminous crops. Nodulation is regulated by a complex network comprising several transcription factors. Here, we functionally characterized the role of a TOC1 family member, GmTOC1b, in soybean (*Glycine max*) nodulation. RT-qPCR assays showed that *GmTOC1b* is constitutively expressed in soybean. However, *GmTOC1b* was also highly expressed in nodules, and GmTOC1 localized to the cell nucleus, based on transient transformation in *Nicotiana benthamiana* leaves. Homozygous *Gmtoc1b* mutant plants exhibited increased root hair curling and produced more infection threads, resulting in more nodules and greater nodule fresh weight. By contrast, *GmTOC1b* overexpression inhibited nodulation. Furthermore, we also showed that GmTOC1b represses the expression of nodulation-related genes including *GmNIN2a* and *GmENOD40-1* by binding to their promoters. We conclude that GmTOC1b functions as a transcriptional repressor to inhibit nodulation by repressing the expression of key nodulation-related genes including *GmNIN2a*, *GmNIN2b*, and *GmENOD40-1* in soybean.

KEYWORDS
soybean, nodulation, *TOC1*, *NIN*, *ENOD40*

Introduction

Soybean (*Glycine max* L.) is an important oil crop and the main source of protein for animal diets. To meet their high nitrogen demand during growth and reproduction, legumes release flavonoids *via* root exudates to attract rhizobia with which they then form nodules (Hungria et al., 2005; Ciampitti and Salvagiotti, 2018). Nodules function as a biological nitrogen fixation factory, which can efficiently convert atmospheric N_2 into ammonia for their plant hosts (Schipanski et al., 2010; Dong et al., 2021b). In exchange, legumes can provide a stable nitrogen fixation environment that is low in oxygen, as well

as carbon and other nutrients for rhizobia. Thus, symbiotic nitrogen fixation not only is a critical factor for improving soybean yield and quality but can also help us reach and maintain sustainable agricultural production practices in an environmentally conscious manner (Gelfand and Philip Robertson, 2015; Santachiara et al., 2019).

The symbiosis between legumes and rhizobia has been well studied. Rhizobia are attracted to plant roots by host-secreted flavonoids and synthesize nodulation (Nod) factors, which are specifically recognized by NOD FACTOR RECEPTOR (NFR) on the root epidermis of legumes (Indrasumunar et al., 2010; Indrasumunar et al., 2011; Liu and Murray, 2016). After initial signal recognition, root hairs curl to surround the rhizobia, followed by the entry of the rhizobia into root endodermis cells through infection lines (Suzaki and Kawaguchi, 2014; Kawaharada et al., 2017). The invaded root cortical cells divide to form a nodule primordium (Heckmann et al., 2011). The rhizobia are finally released into each primordium through cortical cells to develop a mature root nodule (Hayashi et al., 2014).

Many key genes involved in symbiotic nitrogen fixation have been identified. After recognizing Nod factors, NFRs can induce the expression of genes encoding leucine-rich repeat receptor-like kinases (LRR-RLKs) such as *Symbiosis receptor kinase* (*LjSYMRK*) from *Lotus japonicus* (Demchenko et al., 2004), *DOES NOT MAKE INFECTION 2* (*MtDMI2*) from *Medicago truncatula* (Limpens et al., 2005), and *Nodulation receptor kinase* (*GmNORK*) from soybean (Indrasumunar et al., 2015; Wang et al., 2020b), which leads to the activation of downstream gene *DMI3*, encoding the intracellular calcium and calmodulin-dependent protein kinase (CCaMK) (Rival et al., 2012). CCaMK can phosphorylate the transcription factor CYCLOPS and then activate the expression of its downstream target gene *NIN* (*Nodule induction*) (Cerri et al., 2017; Liu et al., 2019). *EARLY NODULIN 40* (*ENOD40*) was the first identified key component of nodulation and encodes a host-derived 12- to 13-amino acid peptide that enhances the stability of sucrose synthase (Crespi et al., 1994; Röhrig et al., 2002). Previous research suggested that *ENOD40* expression can be rapidly induced after inoculation with rhizobia or purified Nod factors, with ENOD40 functioning as an intercellular signal molecule or participating in the regulation of carbon metabolism during nodule formation (Campalans et al., 2004; Xu et al., 2021). In addition, the transcription factors Nodulation Signaling Pathway 1 (NSP1) and NSP2 form a heterologous protein complex in the nucleus (Smit et al., 2005) and directly bind to the promoters of early nodulin genes such as *ENOD11* (Svistoonoff et al., 2010), *NIN* and *Ethylene response factor required for nodule* (*ERN*) (Cerri et al., 2012), which regulate nodule formation at the early stage. Although many studies have investigated nodule signaling, a full picture of nodulation has yet to emerge due to the underlying complex regulatory pathways.

TOC1 (TIMING OF CAB EXPRESSION 1) is a key component of the circadian clock and controls many biological processes supporting plant growth and development (Sanchez et al., 2011; Scheiermann et al., 2013; Xu et al., 2022). In *Arabidopsis* (*Arabidopsis thaliana*), *toc1* mutants enhance drought stress tolerance and showed greater sensitivity to the phytohormone abscisic acid (ABA). Moreover, TOC1 was shown to bind to the promoter of the *ABA receptor* (*ABAR*) and modulate its expression (Legnaioli et al., 2009). TOC1 also interacts with PHYTOCHROME-INTERACTING FACTOR 4 (PIF4) and represses its ability to suppress thermoresponsive growth in the evening (Zhu et al., 2016). In rice (*Oryza sativa*), OsTOC1 regulates tiller-bud and panicle development by indirectly repressing the expression of *TEOSINTE BRANCHED 1* (*TB1*), *DWARF 14* (*D14*), and *IDEAL PLANT ARCHITECTURE 1* (*IPA1*) (Wang et al., 2020a). In the model legume *M. truncatula*, members of other clock components, including LATE ELONGATED HYPOCOTYL (LHY) and LUX ARRHYTHMO (LUX), also regulate nodulation. For instance, MtLHY affected nodulation *via* the regulation of nodule cysteine-rich peptides (Kong et al., 2020; Achom et al., 2021). MtLUX also plays a role in nodule formation, probably through an indirect regulation with MtLHY (Kong et al., 2022).

In this study, we characterized the role of *GmTOC1b* in nodule formation. Knockout mutants of *GmTOC1b* promoted soybean nodulation by affecting hair curling and the number of infection threads. Conversely, overexpression of *GmTOC1b* significantly inhibited nodulation in transgenic hairy roots. *GmTOC1b* regulated multiple nodulation-related genes, including *GmNIN2a*, *GmNIN2b*, and *GmENOD40-1*, in the early stage. Further investigation showed that GmTOC1b represses the expression of *GmNIN2a* and *GmENOD40-1* by binding to their promoters. Taken together, our results revealed the vital role of GmTOC1b in regulating nodule formation in soybean and should be useful for genetic improvement in soybean and other legumes.

Materials and methods

Plant materials and growth conditions

Soybean (*G. max* L.) cultivar Williams 82 (W82) and GUS-tagged *Bradyrhizobium japonicum* strain USDA110 were used in this study. All soybean plant materials were generated in the W82 background.

For tissue-specific expression analysis, soybean seeds were sown and cultured at 26°C under 16-h-light/8-h-dark conditions. Tissue samples were harvested at the indicated growth periods. All samples were frozen in liquid nitrogen and stored at −80°C until total RNA extraction.

For the nodulation phenotyping assay, uniform healthy seeds were surface sterilized with chlorine gas for 14 h and

then germinated in vermiculite under 16-h-light/8-h-dark conditions in a growth room at 26°C. Soybean seedlings were supplied with low-nitrogen nutrient solution containing 150 μM of KNO_3, 120 μM of $Ca(NO_3)_2{\cdot}4H_2O$, 25 μM of $MgCl_2$, 30 μM of $(NH_4)_2SO_4$, 40 μM of Fe-Na-EDTA, 500 μM of $MgSO_4{\cdot}7H_2O$, 1.5 μM of $MnSO_4{\cdot}H_2O$, 1.5 μM of $ZnSO_4{\cdot}7H_2O$, 0.5 μM of $CuSO_4{\cdot}5\ H_2O$, 0.15 μM of $(NH_4)_6Mo_7O_{24}{\cdot}4H_2O$, 2.5 μM of $NaB_4O_7{\cdot}10H_2O$, 500 μM of KH_2PO_4, 540 μM of $CaCl_2$, and 345 μM of K_2SO_4. After unifoliolate leaves were completely opened, the seedlings were inoculated with GUS-tagged *B. japonicum* USDA110 and resuspended in low-nitrogen nutrient solution (OD_{600} = 0.1, 5 ml) at Zeitgeber 12. The nodule phenotype was evaluated 14 days after inoculation (DAI).

Phylogenetic analysis

TOC1 homologous protein sequences from monocotyledons and dicotyledons were aligned using the Clustal X. A neighbor-joining tree was constructed with the software MEGA11 (Tamura et al., 2021). Conserved motifs were extracted with the MEME suite (Bailey et al., 2015).

RNA extraction and RT-qPCR analysis

Total RNA was extracted using a FastPure Plant Total RNA Isolation Kit (Vazyme, Nanjing, China). Removal of genomic DNA and first-strand cDNA synthesis was performed using a PrimeScript RT reagent kit with gDNA Eraser (Takara, Dalian, China). qPCR was conducted on a Roche LightCycle480 system (Roche, Mannheim, Germany) with TB Green (Takara). Three biological replicates were used for all experiments. *GmActin* (*Glyma.02g091900*) was used as the reference gene. All primers used are listed in Supplementary Table 1.

Plasmid construction

CRISPR/Cas9-mediated gene editing was used to generate *Gmtoc1b* mutants. Single-guide RNAs (sgRNA) were predicted using the online tool CRISPR-P 2.0 (http://crispr.hzau.edu.cn/cgi-bin/CRISPR2/CRISPR); the two most reliable sgRNAs targeting different regions of *GmTOC1b* were designed and cloned into the CRISPR/Cas9 vector (Hao et al., 2020). The construct was transformed into *Agrobacterium* (*Agrobacterium tumefaciens*) strain EHA105, which was then used to transform soybean cultivar W82 as previously described (Chen et al., 2018). The genotype of *Gmtoc1b* mutants was verified using Sanger sequencing, and homozygous *Gmtoc1b* mutants were used for further research. For the *GmTOC1b* overexpression construct, the *GmTOC1b* coding sequence without stop codon was amplified and cloned into the *proGmUBI:3×Flag* vector under the control of the soybean *Ubiquitin* (*UBI*) promoter to generate *proGmUBI : GmTOC1b-3×Flag*. The cauliflower mosaic virus (CaMV) 35S promoter was used to drive the expression of green fluorescent protein (*GFP*) for screening positive transformation events.

Subcellular localization of GmTOC1b

The *GmTOC1b* coding sequence without the stop codon was cloned in-frame and upstream of the *GFP* sequence driven by the *GmUBI* promoter. The plasmid was transformed into *Agrobacterium* strain GV3101, which was then transiently infiltrated into *Nicotiana benthamiana* leaves. The plants were grown at 26°C under 16-h-light/8-h-dark conditions for 48 h. Green fluorescent protein (GFP) fluorescence was observed using a Zeiss LSM 800 confocal laser scanning microscope (Zeiss, Oberkochen, Germany). Transiently infiltrated *proGmUBI : GFP N. benthamiana* leaves were used as localization controls. Leaves were stained with 4′,6-diamidino-2-phenylindole (DAPI) to label nuclei.

Soybean hairy root transformation

To obtain transgenic composite plants, soybean hairy root transformation was performed according to a previously published method (Kereszt et al., 2007). Positive transgenic hairy roots were confirmed by detecting GFP fluorescence using a hand-held fluorescence lamp (Luyor 3415RG; Luyor, Shanghai, China). The plants were grown under high-humidity conditions for 1 week to adapt to the environment. For the nodulation phenotyping assay, each plant was inoculated with *B. japonicum* USDA110 (OD_{600} = 0.1, 5 ml). Nodule numbers per root and shoot dry weights were evaluated at 20 DAI.

Observation of root hair curling and infection threads

To detect infection events, 7-day-old seedlings were inoculated with GUS-tagged *B. japonicum* USDA110. Lateral root segments (5 cm in length) from W82 and the *Gmtoc1b* mutants were harvested at 3 and 5 DAI. For the detection of root hair curling, root samples of 3-DAI seedlings were gently washed with sterile water and directly used for observation. For detection of infection threads, 5-DAI root samples were washed and then soaked in X-Gluc solution [100 mM of NaH_2PO_4, 100 mM of Na_2HPO_4, 1 mM of $K_4(Fe(CN)_6)$, 1 mM of $K_3(Fe(CN)_6)$, 1 mg/ml of X-Gluc, 10 mM of Na_2EDTA, 0.1% (v/v) Triton X-100] overnight at 37°C, after which time the stained roots were observed. Digital images were taken with a Zeiss Axio Imager A2 microscope (Zeiss, Oberkochen, Germany).

Transient expression assay

To generate promoter-driven firefly luciferase (*LUC*) constructs, ~2-kb promoter fragments for *NIN2a* and *ENOD40-1* were individually amplified from W82 genomic DNA; the PCR product was purified and cloned into the pGreenII 0800-LUC vector. The resulting *proGmNIN2a:LUC* and *proGmENOD40-1:LUC* constructs were used as reporters, and *proGmUBI:3×Flag* and *proGmUBI : GmTOC1b-3×Flag* were used as effectors. The constructs *proGmUBI : GmTOC1b-3×Flag* and *proGmNIN2a:LUC* or *proGmENOD40-1:LUC* were co-infiltrated into *N. benthamiana* leaves. The constructs *proGmUBI:3×Flag* and *proGmNIN2a-LUC* or *proGmENOD40-1-LUC* were co-infiltrated into *N. benthamiana* leaves as the blank control. Relative Firefly Luciferase; (LUC)/Renilla Luciferase (REN) was analyzed. Three independent biological replicates were performed in this assay. The firefly and Renilla luciferase signals were assayed using a Dual-Luciferase Reporter Assay System (Promega, Madison, WI, USA) and measured on a Biotek Synergy H1 Microplate Reader (Agilent, Santa Clara, CA, USA). Digital images were taken after spraying 1 mM of D-Luciferin sodium salt onto infiltrated *N. benthamiana* leaves (Sangon Biotech, Shanghai, China).

Results

Sequence analysis of GmTOC1b

In this study, we characterized the soybean circadian regulator gene *GmTOC1b* (also known as *PSEUDO-RESPONSE REGULATOR 1* [*PRR1*]), showing high similarity to *Arabidopsis TOC1* (Gan et al., 2020). We downloaded the *GmTOC1b* (*Glyma.06G196200*) sequence from the Phytozome (https://phytozome-next.jgi.doe.gov) database. The cDNA sequence of *GmTOC1b* is 2,597 bp with an open reading frame consisting of 1,677 bp that encodes a 559-amino acid protein. GmTOC1b is predicted to contain an N-terminal pseudoreceiver (PR) domain and a C-terminal [CONSTANS, CONSTANS-LIKE, TOC1] (CCT) domain, which are common to all TOC1 orthologs (Figure 1A). Phylogenetic analysis indicated that GmTOC1b groups into the dicots and is closely related to its soybean TOC1 homologs and other TOC1 proteins from *Arabidopsis*, *Brassica rapa*, and *Solanum lycopersicum* (Figure 1C). We also looked for conserved motifs across TOC1 homologs, which identified the PR (motifs 1, 2, and 3) and CCT (motif 9) domains in all TOC1 proteins (Wang et al., 2020a). Notably, motifs 5 and 8 only existed in the TOC1 proteins from monocots (Figure 1B). These results suggest that TOC1 is relatively well conserved with unique characteristics that have arisen since monocotyledons and dicotyledons diverged.

GmTOC1b is a nuclear protein

The subcellular distribution of GmTOC1b is essential for the prediction of its function. *GmTOC1b-GFP* driven by the *GmUBI* promoter was detectable specifically in the nucleus of *N. benthamiana* mesophyll cells (Figure 2A), whereas the GFP signal was distributed ubiquitously among the *N. benthamiana* mesophyll cells in the *proGmUBI*-GFP control. The results showed that GmTOC1b is a nucleus-localized protein.

Expression patterns of *GmTOC1b* in soybean

We harvested samples from various tissues such as roots, nodules, cotyledons, hypocotyls, stems, primary leaves, trifoliate leaves, the shoot apical meristem (SAM), flowers, pods, and seeds 12 h after lights on (Zeitgeber 12 [ZT12]) from seedlings or plants grown under a 16-h-light/8-h-dark photoperiod. We detected *GmTOC1b* transcripts in all tissues tested, but with the highest abundance in nodules (Figure 2B). Furthermore, we also assessed *GmTOC1b* expression levels in leaves, roots, and nodules of soybean plants 14 days after inoculation (DAI) with *B. japonicum* strain USDA110. We determined that *GmTOC1b* is also highly expressed in nodules (Figure 2C), suggesting that GmTOC1b may function in the formation and development of nodules.

Knockout of *GmTOC1b* promotes soybean nodulation

To investigate a possible role for GmTOC1b in nodulation, we used clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated nuclease 9 (Cas9)-mediated gene editing to generate *Gmtoc1b* mutants. We generated three positive transgenic events from 10 T0 generation transgenic plants. After the screening of heterozygous *Gmtoc1b* mutant lines, we obtained two homozygous T_2 generations *Gmtoc1b* mutants with a 7-bp deletion (*Gmtoc1b-1*) or a 4-bp deletion (*Gmtoc1b-2*) (Supplementary Figure 1A). The T_3 generation of homozygous *Gmtoc1b* mutants was used for further study. Reverse transcription–quantitative PCR (RT-qPCR) analysis showed that *GmTOC1b* expression levels are markedly lower in both *Gmtoc1b* mutants compared to W82 (Supplementary Figure 1B). These two mutants allowed us to investigate GmTOC1b function by evaluating the number of nodules per plant for the *Gmtoc1b* mutants and the wild-type W82 after inoculation with β-glucuronidase (GUS)-tagged *B. japonicum* USDA110. We observed a significant increase in nodule number and nodule

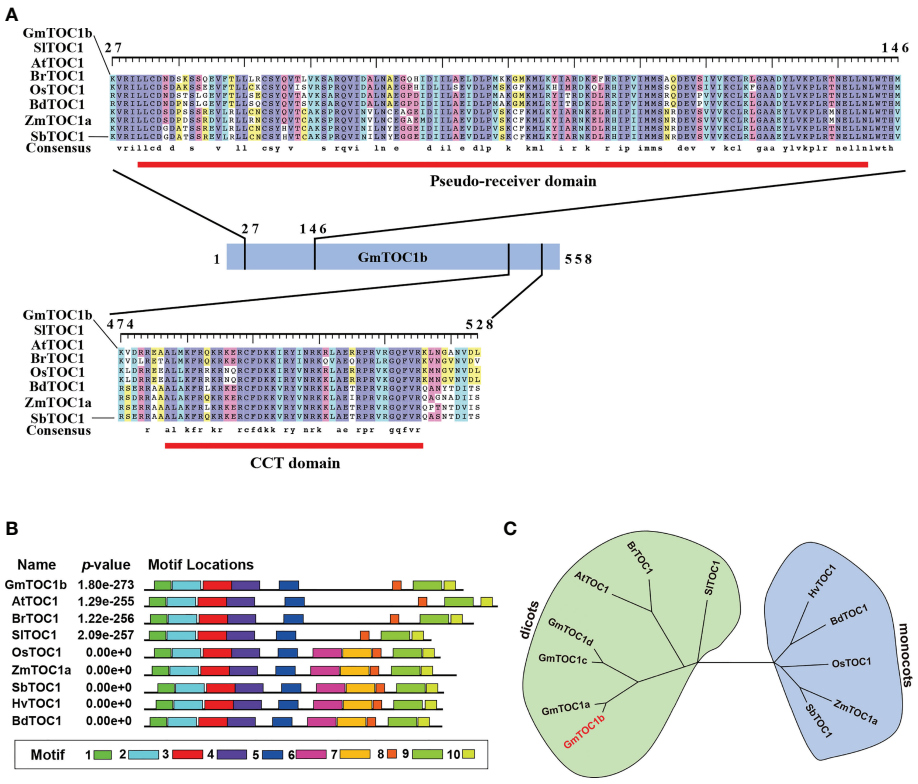


FIGURE 1
Analysis of TOC1 homologs in soybean and other plants. **(A)** TOC1 homologs contain conserved an N-terminal pseudo-receiver domain and a C-terminal CCT (CONSTANS, CO-LIKE, and TOC1) domain. Amino acid sequences were aligned using Clustal X. Red lines indicates the highly conserved domains. aa, amino acids; Sl, *Solanum lycopersicum*; At, *Arabidopsis thaliana*; Br, *Brassica rapa*; Os, *Oryza sativa*; Bd, *Brachypodium distachyon*; Zm, *Zea mays*; Sb, *Sorghum bicolor*. **(B)** Distribution of conserved motifs in TOC1 homologs. The MEME website tool was used to perform the analysis. **(C)** Phylogenetic analysis of TOC1 proteins. A neighbor-joining tree was constructed using full-length protein sequences, with bootstrap values set to 1,000 replicates.

fresh weight in the two *Gmtoc1b* mutants relative to W82 at 14 DAI (Figures 3A–C). As leghemoglobin plays a key role in the biological nitrogen fixation of leguminous plants (Berger et al., 2020; Du et al., 2020; Jiang et al., 2021), we further measured the content of leghemoglobin. The results showed that no significant difference existed between nodules from W82 and *Gmtoc1b* mutants (Supplementary Figure 3). The results indicated that the increased nodule number may not due to the functional weakening of nodules in *Gmtoc1b* mutants.

Knocking out *GmTOC1b* enhances rhizobial infection and nodule formation

To explore the roles of GmTOC1b during rhizobial infection and nodule formation, we observed root phenotypes after inoculation. Compared to W82, the two *Gmtoc1b* mutants both exhibited significantly greater root hair curling at 3 DAI and more infection threads at 5 DAI (Figures 3D–G). Both *Gmtoc1b* mutants also produced more small nodules at 7 DAI relative to W82 (Supplementary Figures 2A, B).

Overexpression of *GmTOC1b* inhibits nodulation

In a complementary approach, we also generated transgenic composite plants with hairy roots overexpressing *GmTOC1b* under the control of the soybean *Ubiquitin* (*GmUBI*) promoter, which we then inoculated with rhizobia. Relative *GmTOC1b* transcript levels were close to 60-fold higher in the overexpression (*GmTOC1b*-OE) lines; moreover, we detected Flag-tagged GmTOC1b, but not in empty vector control transgenic roots (Figures 4B, C). The transgenic *GmTOC1b*-OE roots formed fewer nodules compared to empty vector control hair roots at 20 DAI (Figure 4A). The lower number of nodules per gram of root or shoot dry weight also suggested that overexpression of *GmTOC1b* significantly inhibits

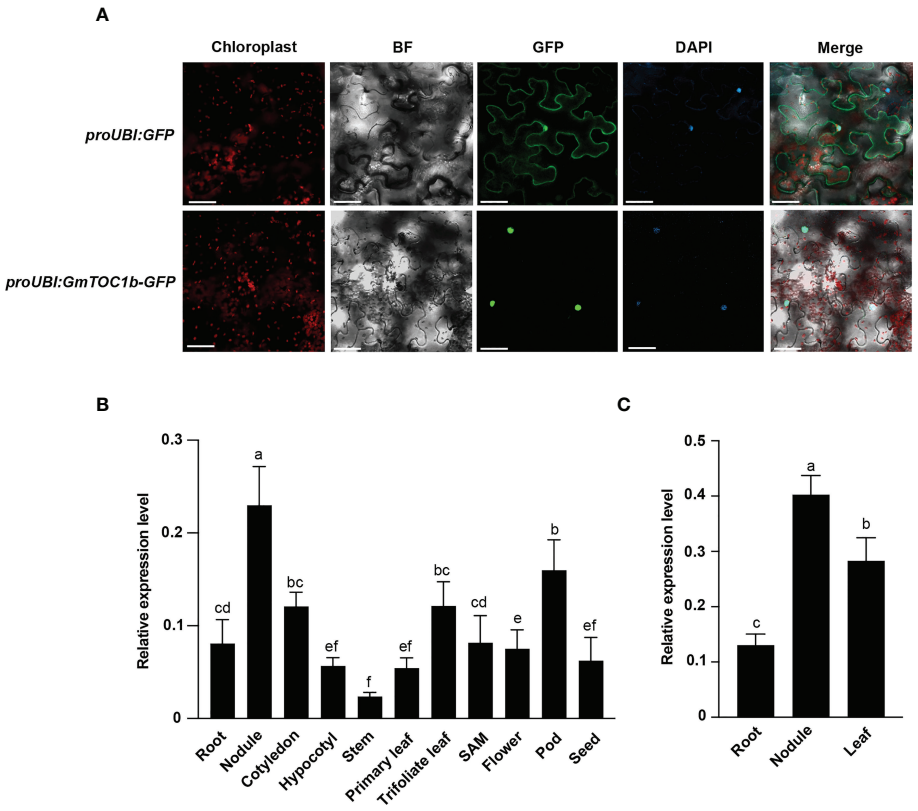


FIGURE 2
GmTOC1b localization and expression pattern analyses of *GmTOC1b*. **(A)** Subcellular localization of GmTOC1b. *Nicotiana benthamiana* leaves were transiently infiltrated with *proUBI : GFP* or *proUBI : GmTOC1b-GFP*; the plants were then grown for 48 h under 16-h-light/8-h-dark condition before GFP fluorescence was observed by microscopy. GFP, green fluorescent protein; BF, brightfield; Merge, GFP and BF images. Scale bars = 20 μm. **(B)** Relative *GmTOC1b* transcript levels in various tissues during growth and at different development stages. Relative expression level was calculated as the ratio of the expression value of the target gene to that of reference gene *GmActin* as an internal standard. **(C)** RT-qPCR analysis of *GmTOC1b* in different tissues at 14 days after inoculation (DAI) with rhizobia. Relative expression level was calculated as the ratio of the expression value of the target gene to that of reference gene *GmActin* as an internal standard. Three biological replicates were performed in all experiments. Different lowercase letters indicate significant differences as determined by Student's *t*-test.

nodulation (Figures 4D, E). Together, these data reveal that GmTOC1b plays a negative role in soybean nodulation.

GmTOC1b inhibits the transcription of nodulation-related genes

To explore the molecular mechanisms by which GmTOC1b inhibits nodule formation, we analyzed the expression pattern of the nodulation-related marker genes *GmNIN2a*, *GmNIN2b*, and *GmENOD40-1* in infected soybean roots at 3 DAI. We first assessed their expression levels in the *Gmtoc1b* mutants, and all three genes were induced in the *Gmtoc1b* mutants (Figures 5A–C). By contrast, these three genes were expressed at much lower levels in transgenic hairy roots overexpressing *GmTOC1b* (Figures 5D–F). These results suggest that GmTOC1b functions as a negative regulator of these nodulation-related genes.

GmTOC1b inhibits *GmNIN2a* and *GmENOD40-1* transcription *via* binding to their promoters

Previous research showed that *Arabidopsis* TOC1 functions as a general transcriptional repressor (Gendron et al., 2012). Based on the lower expression of *GmNIN2a*, *GmNIN2b*, and *GmENOD40-1* in transgenic hairy roots overexpressing *GmTOC1b*, we speculated that GmTOC1b may bind to the promoters of these genes during nodulation. To test this hypothesis, we first analyzed 2-kb fragments of the *GmNIN2a*, *GmNIN2b*, and *GmENOD40-1* promoters to look for *cis*-elements bound by TOC1. We identified two typical HUD motifs in the *GmNIN2a* promoter (Michael et al., 2008a), one typical morning element motif, and three typical HUD motifs in the *GmENOD40-1* promoter (Michael et al., 2008b), suggesting that GmTOC1b may bind to these promoters. We then

10.3389/fpls.2022.1052017

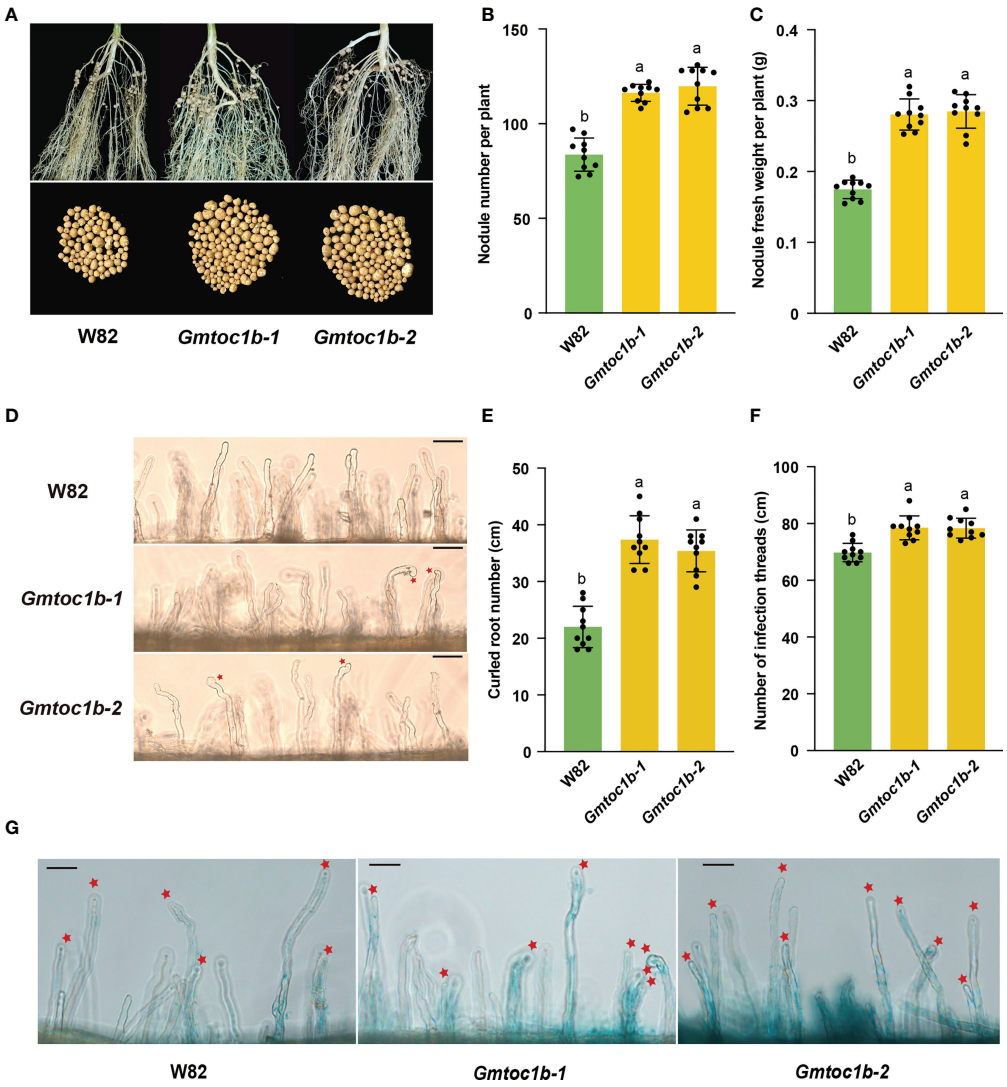


FIGURE 3
Knockout of *GmTOC1b* promotes nodulation in soybean. **(A)** Representative nodule performance of W82 and *Gmtoc1b* mutants at 14 DAI. **(B, C)** Nodule number and nodule fresh weight per plant for W82 and *Gmtoc1b* mutants at 14 DAI. Values are the mean ± SD. **(D)** Representative image of root hair curling in W82 and the *Gmtoc1b* mutant at 3 DAI. Scale bar, 100 μm. **(E)** Number of curled root hairs on W82 and *Gmtoc1b* mutant plants per centimeter of root length (n = 10). **(F)** Number of infection threads in W82 and *Gmtoc1b* mutants per centimeter of root length (n = 10). **(G)** Representative image of infection numbers observed in the W82 and *Gmtoc1b* mutants at 5 DAI (n = 10). Scale bars, 100 μm. Different lowercase letters indicate significant differences as determined by Student's *t*-test.

performed a dual-luciferase transient expression assay using the *GmNIN2a* and *GmENOD40-1* promoters individually driving the firefly luciferase (*LUC*) reporter gene together with an effector construct harboring *GmTOC1b* expressed from the soybean *Ubiquitin* (*GmUBI*) promoter (*proGmUBI : GmTOC1B-3Flag*) (Figure 6A). Co-infiltration of *N. benthamiana* leaves with the effector construct and each LUC reporter construct established that the presence of GmTOC1b significantly inhibits the LUC activity derived from the *GmNIN2a* and *GmENOD40-1* promoters compared to the empty effector vector control (*proGmUBI:3Flag*). Together, these results indicate that GmTOC1b represses the transcription of *GmNIN2a* and *GmENOD40-1*, likely by binding to their promoters (Figures 6B–E).

Discussion

Symbiosis-mediated nitrogen fixation provides a large amount of nitrogen during soybean growth and reproduction (Pagano and Miransari, 2016). Therefore, the effect of nitrogen fixation is particularly important to soybean yield and quality.

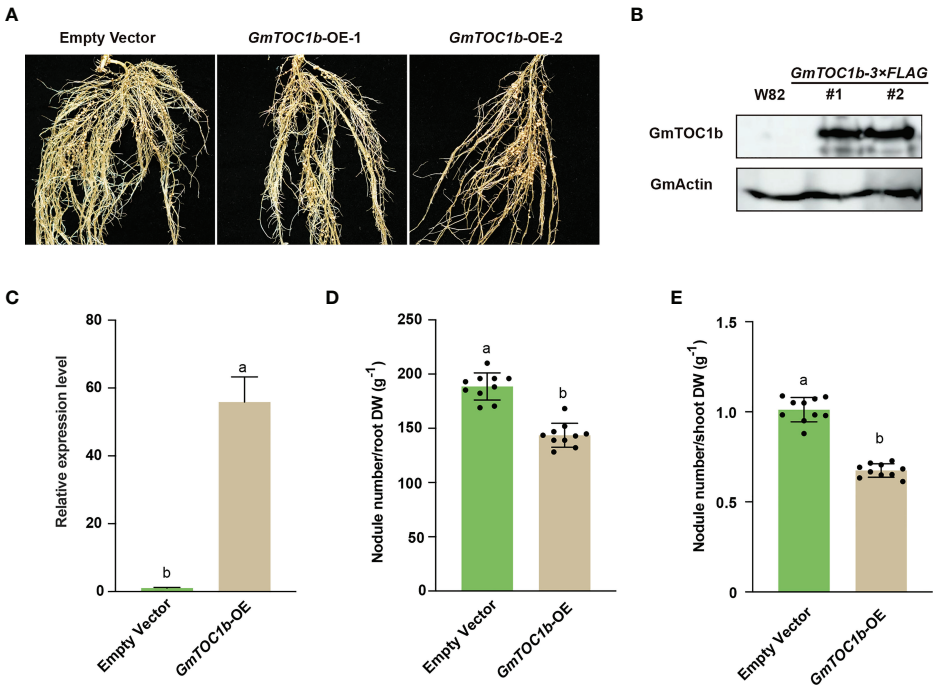


FIGURE 4
Overexpression of *GmTOC1b* in transgenic hairy roots inhibits nodulation. **(A)** Representative nodulation phenotype of *GmTOC1b*-OE and empty vector control transgenic hairy roots at 14 DAI. **(B)** Immunoblot analysis of GmTOC1b abundance in W82 and *GmTOC1b-3Flag* transgenic hairy roots. An anti-Flag antibody was used to recognize GmTOC1b-3Flag; actin was used as loading control. **(C)** Relative *GmTOC1b* transcript levels in *GmTOC1b*-OE and empty vector control transgenic hairy roots. Relative expression level was calculated as the ratio of the expression value of the target gene to that of reference gene *GmActin* as an internal standard. **(D, E)** Nodule number per gram of root dry weight or shoot dry weight in *GmTOC1b*-OE and empty vector control transgenic hairy roots at 14 DAI (n = 10). Different lowercase letters indicate significant differences as determined by Student's *t*-test.

Importantly, the growth and proliferation of rhizobia require host-derived carbohydrates to support optimal utilization of nutrients, driving the co-evolution of a complex and precise regulation network between legumes and their rhizobial symbionts to modulate nodule number. Over the past 10 years, many key regulators of nodule formation have been reported (Kaló et al., 2005; Smit et al., 2005; Lim et al., 2011; Gamas et al., 2017; Suzaki and Nishida, 2019; Wang et al., 2019; Gao et al., 2021; He et al., 2021; Zhang et al., 2021). In soybean, components of the circadian clock, such as *EARLY FLOWERING 3* (*GmELF3*), *LATE ELONGATED HYPOCOTYL* (*GmLHY*), *GmPRR3*, and *LUX ARRHYTHMO* (*GmLUX*), were shown to be involved in the regulation of growth and development or abiotic stress tolerance (Lu et al., 2017; Cheng et al., 2020; Lu et al., 2020; Bu et al., 2021; Dong et al., 2021a; Fang et al., 2021; Li et al., 2021; Wang et al., 2021a). Members of the clock components including LHY and LUX play roles in regulating nodulation in the model legume *M. truncatula*. However, an effect of the circadian clock on nodulation has not previously been reported in soybean. Our results indicated that *GmTOC1b* was highly expressed in nodules. Furthermore, GmTOC1b inhibited soybean nodulation by repressing the transcription of the key nodulation-related genes *GmNIN2a*, *GmNIN2b*, and *GmENOD40-1*, demonstrating that GmTOC1 is a transcriptional repressor.

GmTOC1b is predicted to contain the typical N-terminal PR domain and C-terminal CCT domain of pseudo-response regulators. Previous research indicated that TOC1 directly binds to *cis*-elements located in the promoters of target genes through its CCT domain. Indeed, transient overexpression of *TOC1* in *Arabidopsis* significantly inhibits the transcription of *LHY* and *CIRCADIAN CLOCK-ASSOCIATED 1* (*CCA1*), demonstrating that TOC1 can repress the transcription of its target genes *via* binding to their promoters (Gendron et al., 2012). We confirmed that GmTOC1b and AtTOC1 shared the highly conserved CCT domain, which agreed with the notion that GmTOC1b may bind to the promoters of target genes. In this study, we detected two or more TOC1-binding *cis*-elements in the promoter regions of *GmNIN2a* and *GmENOD40-1*. We further confirmed that GmTOC1b can repress their

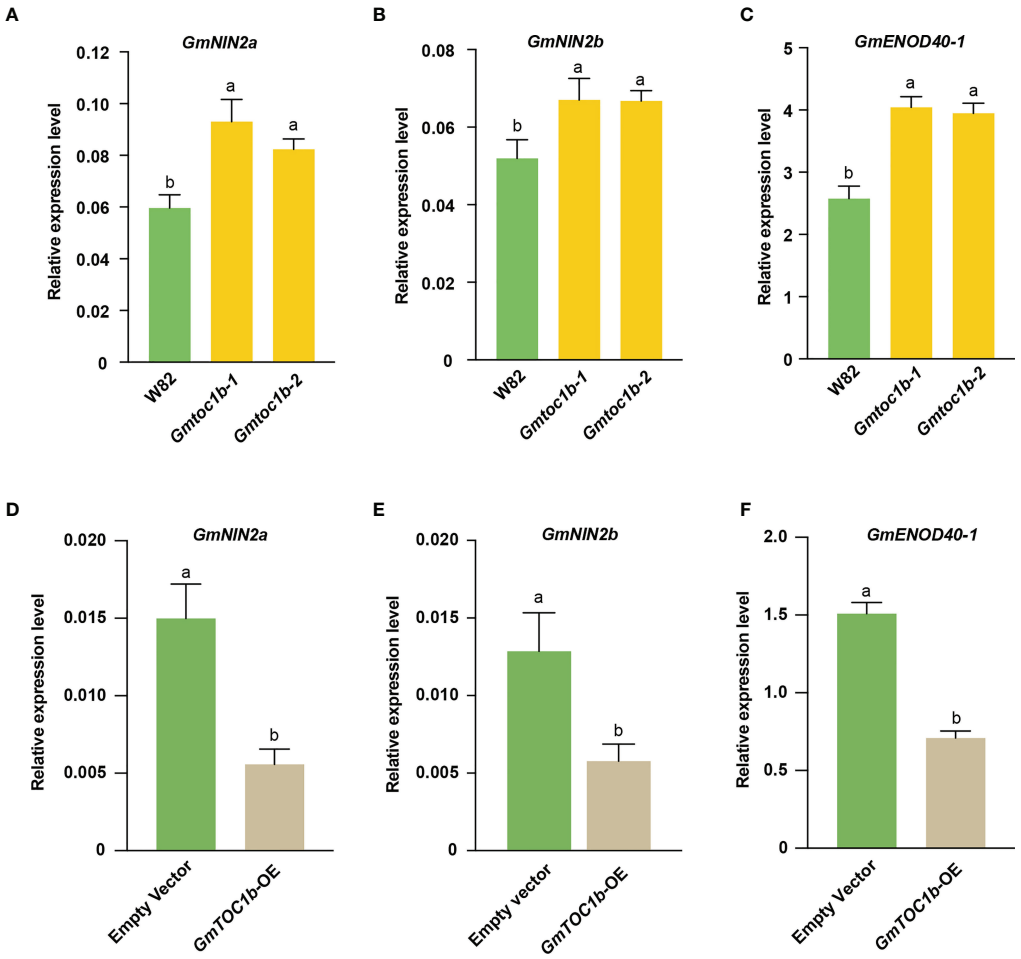


FIGURE 5
Expression of key nodulation-related genes in the early stage of rhizobia inoculation. Roots were harvested for analysis at 3 DAI. **(A–C)** Relative transcript levels of *GmNIN2a*, *GmNIN2b*, and *GmENOD40-1* in the roots of W82 and *Gmtoc1b* mutants. **(D–F)** Relative transcript levels of *GmNIN2a*, *GmNIN2b*, and *GmENOD40-1* in the roots of *GmTOC1b*-OE and empty vector control transgenic hairy roots. Relative expression level was calculated as the ratio of the expression value of the target gene to that of reference gene *GmActin* as an internal standard. Three biological replicates were performed in all experiments. Different lowercase letters indicate significant differences as determined by Student's *t*-test.

transcription, likely by directly binding to the *GmNIN2a* and *GmENOD40-1* promoters (Figures 6B–E).

NINs and ENOD40 play a vital role in rhizobia symbiosis. The expression of their encoding genes is rapidly induced during nodulation, and their corresponding knockout mutants markedly reduced nodule number in both *L. japonicus* (Kumagai et al., 2006; Yoro et al., 2014) and *M. truncatula* (Marsh et al., 2007; Wan et al., 2007). A recent study suggested that light-induced GmSTFs and GmFTs move from shoots to roots, and GmFT2a directly interacts with GmCCaMK-phosphorylated GmSTF3 to form a complex, which directly activates the expression of *NIN* and nuclear factor Y to regulate nodulation (Wang et al., 2021b). Nodule Number Control 1 (GmNNC1) interacts with GmNINa to release the transcriptional repression of *GmENON40* imposed by NNC1 to regulate nodulation (Wang et al., 2019). *SQUAMOSA PROMOTER BINDING PROTEIN-LIKE 9* (*GmSPL9d*) was co-expressed with *GmNINa* and *GmENOD40-1*, and GmSPL9d positively regulated nodulation by inducing the transcription of *GmNINa* and *GmENOD40-1* (Yun et al., 2022). RT-qPCR analysis of the *Gmtoc1b* mutants and *GmTOC1b*-OE transgenic hairy roots showed that the expression levels of *GmNIN2a*, *GmNIN2b*, and *GmENOD40-1* were negatively regulated by GmTOC1b (Figures 5A–F). However, only the transcription of *GmNIN2a* and *GmENOD40-1* was directly repressed by GmTOC1b, while the change in *GmNIN2b* expression could

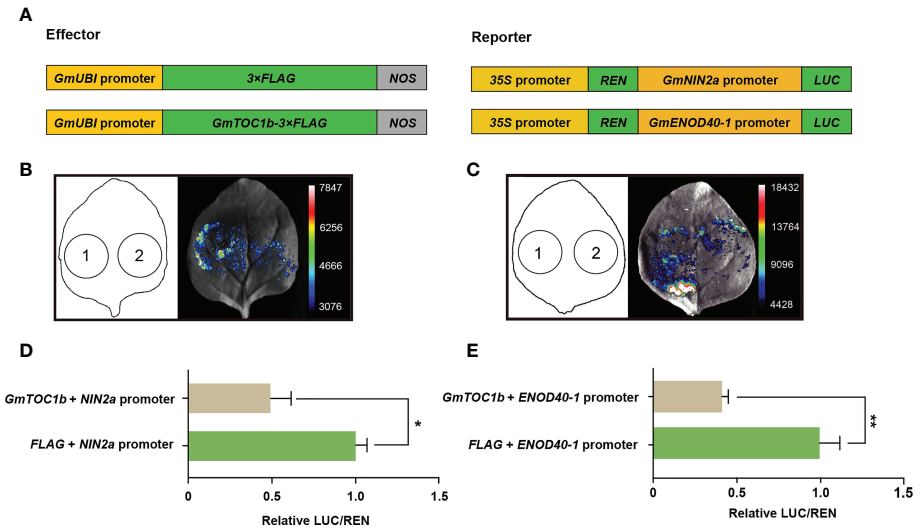


FIGURE 6
GmTOC1b represses the transcription of *GmNIN2a* and *GmENOD40-1*. **(A)** Schematic diagram of the constructs used for the transient co-transfection assay. **(B)** Representative image of firefly luciferase (LUC) activity driven by the *GmNIN2a* promoter. 1, *proGmUBI:3×Flag + proGmNIN2a:LUC*; 2, *proGmUBI : GmTOC1b-3×Flag + proGmNIN2a:LUC*. **(C)** Representative image of luciferase activity driven by the *GmENOD40-1* promoter. 1, *proGmUBI:3×Flag + proGmENOD40-1:LUC*; 2, *proGmUBI : GmTOC1b-3×Flag + proGmENOD40-1:LUC*. **(D, E)** GmTOC1b represses *GmNIN2a* and *GmENOD40-1* promoter activities. Relative firefly luciferase activity driven by the *GmNIN2a* **(D)** or *GmENOD40-1* **(E)** promoter was normalized to CaMV 35S promoter-driven *Renilla* luciferase (REN) activity. Data are means ± SD of seven independent samples (*$p < 0.05$; **$p < 0.01$).

not be explained. Because many genes have been reported to regulate *NIN* expression in soybean, we speculate that GmTOC1b may indirectly regulate *GmNIN2b* expression by directly regulating other nodulation-related genes upstream of *NIN*, which needs to be explored in our future studies.

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

YZ performed phenotypic observation and data analysis. QC performed the gene cloning and generation of *Gmtoc1b* mutant. CL and LL performed the subcellular localization. CG and ZC performed the soybean hairy root transformation. YW performed the transient expression assay. BL, FK, and LC wrote the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the National Natural Science Foundation of China (32001502, 32090064), the Major Program of Guangdong Basic and Applied Research 2019B030302006, and the China Postdoctoral Science Foundation (2020M682655).

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.1052017/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Generation and identification of the *Gmtoc1b* mutants. **(A)** Sequencing verification of CRISPR/Cas9-edited *Gmtoc1b-1* and *Gmtoc1b-2* mutants. **(B)** Relative *GmTOC1b* transcript levels in W82, *Gmtoc1b-1*, and *Gmtoc1b-2* mutants. Three biological replicates were performed in all experiments. All data were normalized to the transcript level of reference gene *GmActin*.

SUPPLEMENTARY FIGURE 2
Knockout of *GmTOC1b* increases small nodule numbers at 7 DAI. **(A)** Representative phenotype of small nodules observed in W82 and the *Gmtoc1b* mutants at 7 DAI. Scale bars, 100 mm. **(B)** Number of small nodules in W82 and the *Gmtoc1b* mutants at 7 DAI (n = 5).

SUPPLEMENTARY FIGURE 3
Content of leghemoglobin in nodules from W82 and *Gmtoc1b* mutants. Data are means ± SD of three independent samples. Lowercase letters indicate significant differences as determined by Student's *t*-test.

References

Achom, M., Roy, P., Lagunas, B., Picot, E., Richards, L., Bonyadi-Pour, R., et al. (2021). Plant circadian clock control of *Medicago truncatula* nodulation *via* regulation of nodule cysteine-rich peptides. *J. Exp. Bot.* 73 (7), 2142–2156. doi: 10.1093/jxb/erab526

Bailey, T. L., Johnson, J., Grant, C. E., and Noble, W. S. (2015). The MEME suite. *Nucleic Acids Res.* 43 (W1), W39–W49. doi: 10.1093/nar/gkv416

Berger, A., Guinand, S., Boscari, A., Puppo, A., and Brouquisse, R. (2020). Medicago truncatula phytoglobin 1.1 controls symbiotic nodulation and nitrogen fixation *via* the regulation of nitric oxide concentration. *New Phytol.* 227, 84–98. doi: 10.1111/nph.16462

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.* 118 (8), e2010241118. doi: 10.1073/pnas.2010241118

Campalans, A., Kondorosi, A., and Crespi, M. (2004). *Enod40*, a short open reading frame–containing mRNA, induces cytoplasmic localization of a nuclear RNA binding protein in *Medicago truncatula*. *Plant Cell* 16 (4), 1047–1059. doi: 10.1105/tpc.019406

Cerri, M. R., Frances, L., Laloum, T., Auriac, M.-C., Niebel, A., Oldroyd, G. E., et al. (2012). *Medicago truncatula* ERN transcription factors: regulatory interplay with NSP1/NSP2 GRAS factors and expression dynamics throughout rhizobial infection. *Plant Physiol.* 160 (4), 2155–2172. doi: 10.1104/pp.112.203190

Cerri, M. R., Wang, Q., Stolz, P., Folgmann, J., Frances, L., Katzer, K., et al. (2017). The ERN1 transcription factor gene is a target of the CCaMK/CYCLOPS complex and controls rhizobial infection in *Lotus japonicus*. *New Phytol.* 215 (1), 323–337. doi: 10.1111/nph.14547

Chen, L., Cai, Y., Liu, X., Yao, W., Guo, C., Sun, S., et al. (2018). Improvement of soybean *Agrobacterium*-mediated transformation efficiency by adding glutamine and asparagine into the culture media. *Int. J. Mol. Sci.* 19 (10), 3039. doi: 10.3390/ijms19103039

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. doi: 10.3389/fpls.2020.00272

Ciampitti, I. A., and Salvagiotti, F. (2018). New insights into soybean biological nitrogen fixation. *Agron. J.* 110 (4), 1185–1196. doi: 10.2134/agronj2017.06.0348

Crespi, M., Jurkevitch, E., Poiret, M., d'Aubenton-Carafa, Y., Petrovics, G., Kondorosi, E., et al. (1994). *Enod40*, a gene expressed during nodule organogenesis, codes for a non-translatable RNA involved in plant growth. *EMBO J.* 13 (21), 5099–5112. doi: 10.1002/j.1460-2075.1994.tb06839.x

Demchenko, K., Winzer, T., Stougaard, J., Parniske, M., and Pawlowski, K. (2004). Distinct roles of *Lotus japonicus SYMRK* and *SYM15* in root colonization and arbuscule formation. *New Phytol.* 163 (2), 381–392. doi: 10.1111/j.1469-8137.2004.01123.x

Dong, L., Fang, C., Cheng, Q., Su, T., Kou, K., Kong, L., et al. (2021a). Genetic basis and adaptation trajectory of soybean from its temperate origin to tropics. *Nat. Commun.* 12 (1), 1–11. doi: 10.1038/s41467-021-25800-3

Dong, W., Zhu, Y., Chang, H., Wang, C., Yang, J., Shi, J., et al. (2021b). An SHR-SCR module specifies legume cortical cell fate to enable nodulation. *Nature* 589 (7843), 586–590. doi: 10.1038/s41586-020-3016-z

Du, M., Gao, Z., Li, X., and Liao, H. (2020). Excess nitrate induces nodule greening and reduces transcript and protein expression levels of soybean leghaemoglobins. *Ann. Bot.* 126 (1), 61–72. doi: 10.1093/aob/mcaa002

Fang, C., Liu, J., Zhang, T., Su, T., Li, S., Cheng, Q., et al. (2021). A recent retrotransposon insertion of *J* caused *E6* locus facilitating soybean adaptation into low latitude. *J. Integr. Plant Biol.* 63 (6), 995–1003. doi: 10.1111/jipb.13034

Gamas, P., Brault, M., Jardinaud, M.-F., and Frugier, F. (2017). Cytokinins in symbiotic nodulation: when, where, what for? *Trends Plant Sci.* 22 (9), 792–802. doi: 10.1016/j.tplants.2017.06.012

Gan, Z., Shi, W., Li, Y., Hou, Z., Li, H., Cheng, Q., et al. (2020). Identification of CRISPR/Cas9 knockout targets and tissue expression analysis of circadian clock genes *GmLNK1/2*, *GmRVE4/8* and *GmTOC1* in soybean. *Acta Agronomica Sin.* 46 (8), 1291–1300. doi: 10.3724/sp.J.1006.2020.94169

Gao, Z., Chen, Z., Cui, Y., Ke, M., Xu, H., Xu, Q., et al. (2021). GmPIN-dependent polar auxin transport is involved in soybean nodule development. *Plant Cell* 33 (9), 2981–3003. doi: 10.1093/plcell/koab183

Gelfand, I., and Philip Robertson, G. (2015). A reassessment of the contribution of soybean biological nitrogen fixation to reactive n in the environment. *Biogeochemistry* 123 (1), 175–184. doi: 10.1007/s10533-014-0061-4

Gendron, J. M., Pruneda-Paz, J. L., Doherty, C. J., Gross, A. M., Kang, S. E., and Kay, S. A. (2012). Arabidopsis circadian clock protein, TOC1, is a DNA-binding transcription factor. *Proc. Natl. Acad. Sci.* 109 (8), 3167–3172. doi: 10.1073/pnas.1200355109

Hao, Y., Zong, W., Zeng, D., Han, J., Chen, S., Tang, J., et al. (2020). Shortened snRNA promoters for efficient CRISPR/Cas-based multiplex genome editing in monocot plants. *Sci. China-Life Sci.* 63, 933–935. doi: 10.1007/s11427-019-1612-6

Hayashi, T., Shimoda, Y., Sato, S., Tabata, S., Imaizumi-Anraku, H., and Hayashi, M. (2014). Rhizobial infection does not require cortical expression of upstream common symbiosis genes responsible for the induction of $Ca^{(2+)}$ spiking. *Plant J.* 77 (1), 146–159. doi: 10.1111/tpj.12374

Heckmann, A. B., Sandal, N., Bek, A. S., Madsen, L. H., Jurkiewicz, A., Nielsen, M. W., et al. (2011). Cytokinin induction of root nodule primordia in *Lotus japonicus* is regulated by a mechanism operating in the root cortex. *Mol. Plant-Microbe Interact.* 24 (11), 1385–1395. doi: 10.1094/MPMI-05-11-0142

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 (3), 488–502. doi: 10.1016/j.molp.2020.12.015

Hungria, M., Franchini, J., Campo, R., and Graham, P. (2005). "The importance of nitrogen fixation to soybean cropping in south America," in *Nitrogen fixation in agriculture, forestry, ecology, and the environment* (Springer), 4, 25–42. doi: 10.1007/1-4020-3544-6_3

Indrasumunar, A., Kereszt, A., Searle, I., Miyagi, M., Li, D., Nguyen, C. D., et al. (2010). Inactivation of duplicated nod factor receptor 5 (NFR5) genes in recessive loss-of-function non-nodulation mutants of allotetraploid soybean (*Glycine max* l. merr.). *Plant Cell Physiol.* 51 (2), 201–214. doi: 10.1093/pcp/pcp178

Indrasumunar, A., Searle, I., Lin, M. H., Kereszt, A., Men, A., Carroll, B. J., et al. (2011). Nodulation factor receptor kinase 1α controls nodule organ number in soybean (*Glycine max* l. merr). *Plant J.* 65 (1), 39–50. doi: 10.1111/j.1365-313X.2010.04398.x

Indrasumunar, A., Wilde, J., Hayashi, S., Li, D., and Gresshoff, P. M. (2015). Functional analysis of duplicated symbiosis receptor kinase (SymRK) genes during nodulation and mycorrhizal infection in soybean (Glycine max). *J. Plant Physiol.* 176, 157–168. doi: 10.1016/j.jplph.2015.01.002

Jiang, S., Jardinaud, M., GAO, J., Pecrix, Y., Wen, J., Mysore, K., et al. (2021). NIN-like protein transcription factors regulate leghemoglobin genes in legume nodules. *Science* 374 (6567), 625–628. doi: 10.1126/science.abg5945

Kaló, P., Gleason, C., Edwards, A., Marsh, J., Mitra, R. M., Hirsch, S., et al. (2005). Nodulation signaling in legumes requires NSP2, a member of the GRAS family of transcriptional regulators. *Science* 308 (5729), 1786–1789. doi: 10.1126/science.1110951

Kawaharada, Y., James, E. K., Kelly, S., Sandal, N., and Stougaard, J. (2017). The ethylene responsive factor required for nodulation 1 (ERN1) transcription factor is required for infection-thread formation in *Lotus japonicus*. *Mol. Plant-Microbe Interact.* 30 (3), 194–204. doi: 10.1094/MPMI-11-16-0237-R

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 (4), 948–952. doi: 10.1038/nprot.2007.141

Kong, Y., Han, L., Liu, X., Wang, H., Wen, L., Yu, X., et al. (2020). The nodulation and nyctinastic leaf movement is orchestrated by clock gene *LHY* in *Medicago truncatula*. *J. Integr. Plant Biol.* 62 (12), 1880–1895. doi: 10.1111/jipb.12999

Kong, Y., Zhang, Y., Liu, X., Meng, Z., Yu, X., Zhou, C., et al. (2022). The conserved and specific roles of the LUX arrhythmo in circadian clock and nodulation. *Int. J. Mol. Sci.* 23 (7), 3473. doi: 10.3390/ijms23073473

Kumagai, H., Kinoshita, E., Ridge, R. W., and Kouchi, H. (2006). RNAi knockdown of *ENOD40s* leads to significant suppression of nodule formation in *Lotus japonicus*. *Plant Cell Physiol.* 47 (8), 1102–1111. doi: 10.1093/pcp/pcj081

Legnaioli, T., Cuevas, J., and Mas, P. (2009). TOC1 functions as a molecular switch connecting the circadian clock with plant responses to drought. *EMBO J.* 28 (23), 3745–3757. doi: 10.1038/emboj.2009.297

Li, Z., Cheng, Q., Gan, Z., Hou, Z., Zhang, Y., Li, Y., et al. (2021). Multiplex CRISPR/Cas9-mediated knockout of soybean *LNK2* advances flowering time. *Crop J.* 9 (4), 767–776. doi: 10.1016/j.cj.2020.09.005

Lim, C. W., Lee, Y. W., and Hwang, C. H. (2011). Soybean nodule-enhanced CLE peptides in roots act as signals in GmNARK-mediated nodulation suppression. *Plant Cell Physiol.* 52 (9), 1613–1627. doi: 10.1093/pcp/pcr091

Limpens, E., Mirabella, R., Fedorova, E., Franken, C., Franssen, H., Bisseling, T., et al. (2005). Formation of organelle-like N_2-fixing symbiosomes in legume root nodules is controlled by DMI2. *Proc. Natl. Acad. Sci.* 102 (29), 10375–10380. doi: 10.1073/pnas.0504284102

Liu, C. W., and Murray, J. D. (2016). The role of flavonoids in nodulation host-range specificity: an update. *Plants (Basel)* 5 (3), 33. doi: 10.3390/plants5030033

Liu, J., Rutten, L., Limpens, E., van der Molen, T., Van Velzen, R., Chen, R., et al. (2019). A remote cis-regulatory region is required for *NIN* expression in the pericycle to initiate nodule primordium formation in *Medicago truncatula*. *Plant Cell* 31 (1), 68–83. doi: 10.1105/tpc.18.00478

Lu, S., Dong, L., Fang, C., Liu, S., Kong, L., Cheng, Q., et al. (2020). Stepwise selection on homeologous *PRR* genes controlling flowering and maturity during soybean domestication. *Nat. Genet.* 52 (4), 428–436. doi: 10.1038/s41588-020-0604-7

Lu, S., Zhao, X., Hu, Y., Liu, S., Nan, H., Li, X., et al. (2017). Natural variation at the soybean *J* locus improves adaptation to the tropics and enhances yield. *Nat. Genet.* 49 (5), 773. doi: 10.1038/ng.3819

Marsh, J. F., Rakocevic, A., Mitra, R. M., Brocard, L., Sun, J., Eschstruth, A., et al. (2007). *Medicago truncatula* NIN is essential for rhizobial-independent nodule organogenesis induced by autoactive calcium/calmodulin-dependent protein kinase. *Plant Physiol.* 144 (1), 324–335. doi: 10.1104/pp.106.093021

Michael, T. P., Breton, G., Hazen, S. P., Priest, H., Mockler, T. C., Kay, S. A., et al. (2008a). A morning-specific phytohormone gene expression program underlying rhythmic plant growth. *PloS Biol.* 6 (9), e225. doi: 10.1371/journal.pbio.0060225

Michael, T. P., Mockler, T. C., Breton, G., McEntee, C., Byer, A., Trout, J. D., et al. (2008b). Network discovery pipeline elucidates conserved time-of-day-specific cis-regulatory modules. *PloS Genet.* 4 (2), e14. doi: 10.1371/journal.pgen.0040014

Pagano, M. C., and Miransari, M. (2016). "The importance of soybean production worldwide," in *Abiotic and biotic stresses in soybean production* (ACADEMIC PRESS), 1–26. doi: 10.1016/B978-0-12-801536-0.00001-3

Rival, P., De Billy, F., Bono, J.-J., Gough, C., Rosenberg, C., and Bensmihen, S. (2012). Epidermal and cortical roles of NFP and DMI3 in coordinating early steps of nodulation in *Medicago truncatula*. *Development* 139 (18), 3383–3391. doi: 10.1242/dev.081620

Röhrig, H., Schmidt, J., Miklashevichs, E., Schell, J., and John, M. (2002). Soybean ENOD40 encodes two peptides that bind to sucrose synthase. *Proc. Natl. Acad. Sci.* 99 (4), 1915–1920. doi: 10.1073/pnas.022664799

Sanchez, A., Shin, J., and Davis, S. J. (2011). Abiotic stress and the plant circadian clock. *Plant Signaling Behav.* 6 (2), 223–231. doi: 10.4161/psb.6.2.14893

Santachiara, G., Salvagiotti, F., and Rotundo, J. L. (2019). Nutritional and environmental effects on biological nitrogen fixation in soybean: A meta-analysis. *Field Crops Res.* 240, 106–115. doi: 10.1016/j.fcr.2019.05.006

Scheiermann, C., Kunisaki, Y., and Frenette, P. S. (2013). Circadian control of the immune system. *Nat. Rev. Immunol.* 13 (3), 190–198. doi: 10.1038/nri3386

Schipanski, M., Drinkwater, L., and Russelle, M. (2010). Understanding the variability in soybean nitrogen fixation across agroecosystems. *Plant Soil* 329 (1), 379–397. doi: 10.1007/s11104-009-0165-0

Smit, P., Raedts, J., Portyanko, V., Debellé, F., Gough, C., Bisseling, T., et al. (2005). NSP1 of the GRAS protein family is essential for rhizobial nod factor-induced transcription. *Science* 308 (5729), 1789–1791. doi: 10.1126/science.1111025

Suzaki, T., and Kawaguchi, M. (2014). Root nodulation: a developmental program involving cell fate conversion triggered by symbiotic bacterial infection. *Curr. Opin. Plant Biol.* 21, 16–22. doi: 10.1016/j.pbi.2014.06.002

Suzaki, T., and Nishida, H. (2019). Autoregulation of legume nodulation by sophisticated transcriptional regulatory networks. *Mol. Plant* 12 (9), 1179–1181. doi: 10.1016/j.molp.2019.07.008

Svistoonoff, S., Sy, M.-O., Diagne, N., Barker, D. G., Bogusz, D., and Franche, C. (2010). Infection-specific activation of the *Medicago truncatula ENOD11* early nodulin gene promoter during actinorhizal root nodulation. *Mol. Plant-Microbe Interact.* 23 (6), 740–747. doi: 10.1094/MPMI-23-6-0740

Tamura, K., Stecher, G., and Kumar, S. (2021). MEGA11: Molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.* 38 (7), 3022–3027. doi: 10.1093/molbev/msab120

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 (5), 2660–2675. doi: 10.1111/nph.17019

Wang, L., Deng, L., Bai, X., Jiao, Y., Cao, Y., and Wu, Y. (2020b). Regulation of nodule number by GmNORK is dependent on expression of *GmNIC* in soybean. *Agroforestry Syst.* 94 (1), 221–230. doi: 10.1007/s10457-019-00382-8

Wang, T., Guo, J., Peng, Y., Lyu, X., Liu, B., Sun, S., et al. (2021b). Light-induced mobile factors from shoots regulate rhizobium-triggered soybean root nodulation. *Science* 374 (6563), 65–71. doi: 10.1126/science.abh2890

Wang, F., Han, T., Song, Q., Ye, W., Song, X., Chu, J., et al. (2020a). The rice circadian clock regulates tiller growth and panicle development through strigolactone signaling and sugar sensing. *Plant Cell* 32 (10), 3124–3138. doi: 10.1105/tpc.20.00289

Wang, L., Sun, Z., Su, C., Wang, Y., Yan, Q., Chen, J., et al. (2019). A GmNINa-miR172c-NNC1 regulatory network coordinates the nodulation and autoregulation of nodulation pathways in soybean. *Mol. Plant* 12 (9), 1211–1226. doi: 10.1016/j.molp.2019.06.002

Wan, X., Hontelez, J., Lillo, A., Guarnerio, C., van de Peut, D., Fedorova, E., et al. (2007). *Medicago truncatula* ENOD40-1 and ENOD40-2 are both involved in nodule initiation and bacteroid development. *J. Exp. Bot.* 58 (8), 2033–2041. doi: 10.1093/jxb/erm072

Xu, H., Li, Y., Zhang, K., Li, M., Fu, S., Tian, Y., et al. (2021). miR169c-NFYA-C-ENOD40 modulates nitrogen inhibitory effects in soybean nodulation. *New Phytol.* 229 (6), 3377–3392. doi: 10.1111/nph.17115

Xu, X., Yuan, L., Yang, X., Zhang, X., Wang, L., and Xie, Q. (2022). Circadian clock in plants: Linking timing to fitness. *J. Integr. Plant Biol.* 64 (4), 792–811. doi: 10.1111/jipb.13230

Yoro, E., Suzaki, T., Toyokura, K., Miyazawa, H., Fukaki, H., and Kawaguchi, M. (2014). A positive regulator of nodule organogenesis, NODULE INCEPTION, acts as a negative regulator of rhizobial infection in *Lotus japonicus*. *Plant Physiol.* 165 (2), 747–758. doi: 10.1104/pp.113.233379

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 (4), 1881–1899. doi: 10.1111/nph.17899

Zhang, B., Wang, M., Sun, Y., Zhao, P., Liu, C., Qing, K., et al. (2021). *Glycine max NNL1* restricts symbiotic compatibility with widely distributed bradyrhizobia *via* root hair infection. *Nat. Plants* 7 (1), 73–86. doi: 10.1038/s41477-020-00832-7

Zhu, J.-Y., Oh, E., Wang, T., and Wang, Z.-Y. (2016). TOC1–PIF4 interaction mediates the circadian gating of thermoresponsive growth in arabidopsis. *Nat. Commun.* 7 (1), 13692. doi: 10.1038/ncomms13692

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 17 November 2022
DOI 10.3389/fpls.2022.1064685

OPEN ACCESS

EDITED BY
Zhengjun Xia,
Chinese Academy of Sciences (CAS), China

REVIEWED BY
Zhenfeng Jiang,
Northeast Agricultural University, China
Zhijuan Wang,
Huazhong Agricultural University, China
Akito Kaga,
Institute of Crop Science (NARO), Japan

*CORRESPONDENCE
Zhenguo Shen
zgshen@njau.edu.cn
Xin Chen
cx@jaas.ac.cn

SPECIALTY SECTION
This article was submitted to Functional and Applied Plant Genomics, a section of the journal Frontiers in Plant Science

RECEIVED 08 October 2022
ACCEPTED 01 November 2022
PUBLISHED 17 November 2022

CITATION
Li X, Jia Y, Sun M, Ji Z, Zhang H, Qiu D, Cai Q, Xia Y, Yuan X, Chen X and Shen Z (2022) *MINI BODY1*, encoding a MATE/DTX family transporter, affects plant architecture in mungbean (*Vigna radiata* L.). *Front. Plant Sci.* 13:1064685. doi: 10.3389/fpls.2022.1064685

MINI BODY1, encoding a MATE/DTX family transporter, affects plant architecture in mungbean (*Vigna radiata* L.)

Xin Li[1], Yahui Jia[1], Mingzhu Sun[1], Zikun Ji[2], Hui Zhang[3], Dan Qiu[1], Qiao Cai[1], Yan Xia[1], Xingxing Yuan[4], Xin Chen[4]* and Zhenguo Shen[1,5]*

[1]College of Life Sciences, Nanjing Agricultural University, Nanjing, China, [2]College of Agro-Grassland Science, Nanjing Agricultural University, Nanjing, China, [3]National experimental Teaching Center for Plant Production, Nanjing Agricultural University, Nanjing, China, [4]Institute of Industrial Crops, Jiangsu Academy of Agricultural Sciences, Nanjing, Jiangsu, China, [5]Jiangsu Collaborative Innovation Center for Solid Organic Waste Resource Utilization, Nanjing Agricultural University, Nanjing, China

It has been shown that multidrug and toxic compound extrusion/detoxification (MATE/DTX) family transporters are involved in the regulation of plant development and stress response. Here, we characterized the *mini body1* (*mib1*) mutants in mungbean, which gave rise to increased branches, pentafoliate compound leaves, and shortened pods. Map-based cloning revealed that *MIB1* encoded a MATE/DTX family protein in mungbean. qRT-PCR analysis showed that *MIB1* was expressed in all tissues of mungbean, with the highest expression level in the young inflorescence. Complementation assays in *Escherichia coli* revealed that MIB1 potentially acted as a MATE/DTX transporter in mungbean. It was found that overexpression of the *MIB1* gene partially rescued the shortened pod phenotype of the *Arabidopsis dtx54* mutant. Transcriptomic analysis of the shoot buds and young pods revealed that the expression levels of several genes involved in the phytohormone pathway and developmental regulators were altered in the *mib1* mutants. Our results suggested that *MIB1* plays a key role in the control of plant architecture establishment in mungbean.

KEYWORDS

legume, mungbean, plant architecture, MIB1, MATE/DTX family, RNA-Seq

Introduction

Plant architecture refers to the three-dimensional organization of plant organs, including the branching pattern and the shape and size of lateral organs, which affects plant growth and productivity (Reinhardt and Kuhlemeier, 2002; Wang and Li, 2008). During the last decades, multiple regulators in the control of plant architecture have been

identified in model plants, such as rice (*Oryza sativa*) and *Arabidopsis thaliana*, which form complex regulatory networks including microRNA, key transcription factors, and phytohormones (Wang and Li, 2008; Guo et al., 2020).

The multidrug and toxic compound extrusion/detoxification (MATE/DTX) family was one of the important groups of multidrug transporters, which plays diverse roles in stress responses including detoxification, iron homeostasis, and drought stress (Diener et al., 2001; Li et al., 2002; Nawrath et al., 2002; Rogers and Guerinot, 2002; Magalhaes et al., 2007; Ishihara et al., 2008; Lu et al., 2019; Upadhyay et al., 2019; Duan et al., 2022; Nimmy et al., 2022). MATE/DTX family proteins also participate in plant development and growth (Thompson et al., 2010; Burko et al., 2011; Li et al., 2014; Suzuki et al., 2015; Jia et al., 2019; Upadhyay et al., 2020; Gani et al., 2022). For example, *Arabidopsis* ADP1/DTX51, a putative MATE/DTX family transporter, affects plant architecture. Elevated expression of *ADP1/DTX51* in *Arabidopsis* leads to an increase in plant growth rate and branch number by modulating the auxin level (Li et al., 2014). Another MATE/DTX transporter, BIG EMBRYO1 (BIGE1) in maize, regulates embryo development, initiation, and the size of lateral organs (Suzuki et al., 2015). The mutation of the maize *BIGE1* gene results in increased leaf number and larger embryo size. Similarly, the mutant of *DTX54/BIGE1A* (ortholog of *BIGE1* in *Arabidopsis*) exhibits increased leaf number and shortened pods with smaller seeds (Suzuki et al., 2015).

Legume is the third largest plant family, with more than 600 genus and 18,000 species (Graham and Vance, 2003). The plant architecture significantly affects the seed yield of grain legume such as pea (*Pisum sativa*), soybean (*Glycine max*), and mungbean (*Vigna radiata*). In pea, the TCP family gene *PsBRC1* integrates phytohormones including auxin, cytokinin (CK) and strigolactones (SL) to regulate shoot branching (Rameau et al., 2015; Kerr et al., 2021). It has been shown that the soybean gene *INCREASED LEAF PETIOLE ANGLE 1* (*GmILPA1*), encoding a subunit of the anaphase-promoting complex, controls the angle of leaf petiole (Gao et al., 2017). Notably, the *MicroRNA156* (*miR156*)-*SQUAMOSA PROMOTER BINDING PROTEIN-LIKE* (*SPL*) module has important roles in controlling plant architecture and agronomic traits in soybean (Bao et al., 2019; Sun et al., 2019). Overexpression of the *GmmiR156b* in soybean significantly alters plant architecture and improves seed yield (Sun et al., 2019). Consistently, knockout *GmmiR156b* targeted gene *GmSPL9* by gene editing alters plant architecture with improved performance and productivity in soybean (Bao et al., 2019). Recently, it has been shown that an MYB family transcription factor GmMYB14 in soybean regulates plant architecture through the brassinosteroid pathway. GmMYB14-overexpressing soybean plants display the compact plant architecture and improved seed yield (Chen et al., 2021). However, up to now, only a few key factors regulating plant architecture has been identified in legume and the underlying molecular mechanism is still poorly understood (Liu et al., 2020).

In this study, we characterized the *mini body1* (*mib1*) mutant in mungbean, which affected plant growth rate, branch number, and lateral organ size. It was found that *MIB1* encoded a member of MATE/DTX family proteins, potentially acting as a transporter in mungbean. Transcriptomic analysis revealed that expression levels of phytohormone pathway genes and developmental regulators were altered in the *mib1* mutants. Our results indicated that MIB1 plays a pivotal role in regulating plant architecture in mungbean.

Materials and methods

Plant materials

Three mutants, namely, *mib1-1* (A001), *mib1-2* (A006), and *mib1-3* (I007), were identified from M_2 generation of the gamma ray mutagenized cultivar Sulu (Li et al., 2022). For phenotype analysis of wild-type (WT) plants, mutants were grown in the greenhouse at 28 ± 2°C, with a 16-h/8-h day/night photoperiod. The allelic tests for three mutants were carried out by crossing the *mib1-1* mutant with the *mib1-2* and *mib1-3* mutants, respectively. All plants of F_1 generation showed the mutated phenotype.

Scanning electron microscopy analysis

The terminal leaflets of the fifth compound leaves were fixed in FAA solution and then the samples were dehydrated in the ethanol/tert-butanol series. Field emission scanning electron microscopic (SU8010, Hitachi, Tokyo, Japan) analysis was conducted as previously described (Jiao et al., 2019).

Map-based cloning of *MIB1* gene

The *mib1-3* mutants were crossed with cultivar AL127 to generate a population for genetic mapping. A total of 150 plants with mutant phenotype isolated from 642 plants in the F_2 population were used to map the *MIB1* gene. The primers of the molecular markers used in present study are listed in Supplementary Table 1. The DNA were extracted *via* a plant Genomic DNA Kit DP305 (Tiangen, Beijing, China). The polymerase chain reaction (PCR) was carried out and the polymorphisms of the markers were analyzed as previously described (Jiao et al., 2016).

The PCR of the *MIB1* genomic region was conducted by the primers in Supplementary Table 1. The PCR products were cloned into the pMD18-T (TaKaRa, Dalian, China) and sequenced.

RNA-sequencing analysis and quantitative reverse transcription PCR analysis

Shoot buds (2 weeks after germination) and the young pods (2 days after pollination) of WT and *mib1-3* mutants were collected with three biological replicates. RNA was extracted by the RNA Kit R6827-01 (Omega, Shanghai, China). We performed RNA-seq using the Illumina HiSeq X Ten platform (Illumina, San Diego, California, USA). The raw sequences were submitted to the NCBI SRA database with accession numbers SRR16944233–SRR16944244. Number of reads per kilobase of exon region in a gene per million mapped reads (RPKM) was used to value expression levels (Mortazavi et al., 2008), and VC1973A version 1.0 was used as the reference genome (Kang et al., 2014). Based on the methods described by Audic and Claverie (1997), DEGs were identified. Heat maps were generated by the pheatmap package (https://cran.r-project.org).

For qRT-PCR, the first strand cDNA was synthesized *via* Takara PrimeScript™ RT reagent Kit RR047A (TaKaRa, Dalian, China). qRT-PCR analysis was conducted using TB Green™ Premix Ex™ RR420A (TaKaRa) and the ABI StepOnePlus machine (Applied Biosystems, Foster City, CA, USA). Three biological replicates with three technical repeats were conducted.

Arabidopsis transformation

The WT (Col-0) and *dtx54* mutant (WiscDsLoxHs046_04F) were used in the present study. The CDS of the *MIB1* gene was cloned into pCAMBIA1304 using primers in Supplementary Table 1. The construct was transformed into the *dtx54* mutants through floral dip transformation as previously described (Clough and Bent, 1998). T_3 progeny lines of *35S::MIB1/dtx54* (L04 and L06) were used for phenotype analysis in this study.

Complementation assays in *Escherichia coli*

WT strain K12 and *ΔacrB* mutant strain of *E. coli* were obtained from Professor Chuanzhen Jiang (South China Agricultural University). The CDS of the mungbean *MIB1* gene was cloned into the pET32a vector using primers in Supplementary Table 1, and the vectors were transformed into K12 and mutant strain. Transformants were selected on Luria-Bertani (LB) plate medium with 100 μg/ml ampicillin. The positive clones were then grown in liquid medium containing ampicillin and 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) to induce the expression of *MIB1*. The cells were diluted and spotted on medium plates with or without tetrabutyl ammonium (TBA) at 37°C for 24 h. Cell growth curves were determined by the absorbance at 600 nm of the cultures grown at 37°C for 24 h.

Analysis of indole-3-acetic acid and abscisic acid contents

Plant hormone levels of indole-3-acetic acid (IAA) and abscisic acid (ABA) in young pods of the WT plant and mutants were determined by high-performance liquid chromatography–mass spectrum/mass spectrum (HPLC/MS/MS) by Agilent 1290 HPLC (Agilent, Santa Clara, CA, USA) and SCIEX-6500 Qtrap (AB Sciex, Foster, CA, USA), as described previously (Pan et al., 2010).

Phylogenetic analysis

In this study, the MIB1 protein sequence was used to search against the mungbean database (Kang et al., 2014). The phylogenetic analysis was conducted by MEGA (version 7.0) using the neighbor-joining method with 1,000 replications (Kumar et al., 2016). The tree was displayed by the Interactive Tree of Life (iTOL; Letunic and Bork, 2016). Protein sequences from this study are listed in Supplementary Table 2.

Results

Isolation and characterization of the *mib1* mutants in mungbean

To investigate key components regulating plant architecture in mungbean, we screened mutants with altered branch number and shape and size of lateral organs from the mutagenesis population (Li et al., 2022). Three allelic mutants affecting plant architecture were isolated in mungbean (Figure 1). We named these mutants *mini body1-1* (*mib1-1*), *mib1-2*, and *mib1-3*, respectively.

The leaf production rate in the *mib1* mutants was accelerated, compared with that of WT (Figures 1A, B). The juvenile leaves of the mutants were normal, but the adult leaves displayed pentafoliate form, compared to those of WT with trifoliate compound leaves (Figure 1C). In the *mib1* mutants, the size of the leaflets was severely reduced by 43.61%–60.93% (Figure 1D). The outgrowth of axillary buds in the *mib1* mutants was faster than those in WT (Figure 1E). The number of branches in the *mib1* mutants increased significantly (Figure 1F). At 4 weeks after germination, there was only one branch in each WT plant, while each *mib1* mutant had four branches (Figure 1F). At 8 weeks after germination, there was no difference in the number of primary branches between WT and

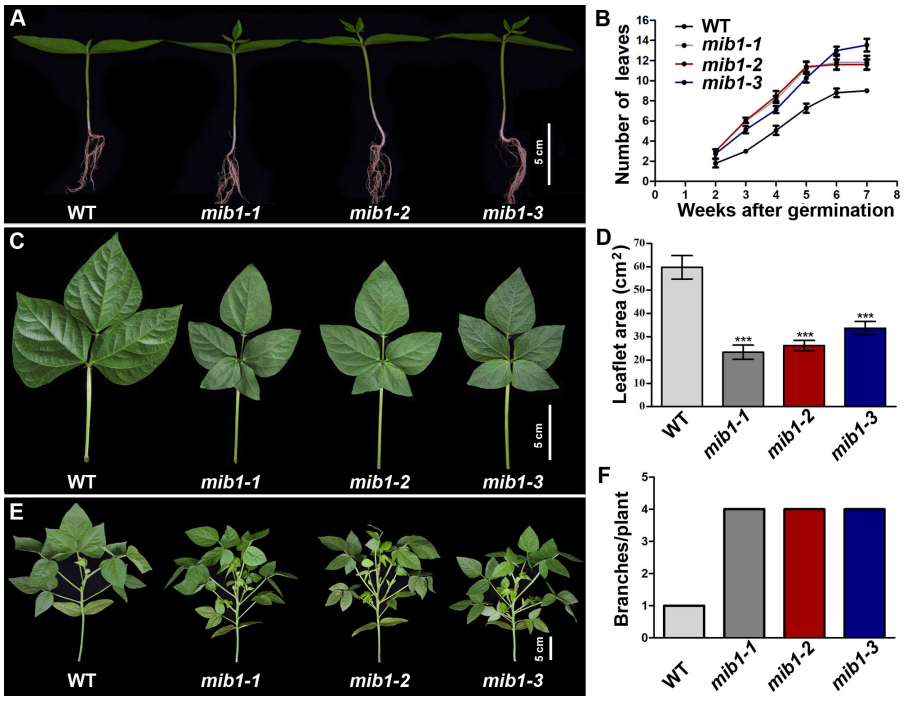


FIGURE 1
Growth rate characterization of WT and *mib1* mutants. **(A)** Two-week-old seedlings of the wild-type (WT) plant and *mib1* mutants. **(B)** The number of compound leaves of WT and *mib1* mutants (n = 10). **(C)** The fifth compound leaves of WT and *mib1* mutants. **(D)** The size of the terminal leaflets of the fifth compound leaves of WT and *mib1* mutants (n = 10). **(E)** Plant architecture of WT and *mib1* mutants at the 4 weeks after germination. **(F)** The number of branches of WT and *mib1* mutants at 4 weeks after germination (n = 10). The data were means ± SD. The Tukey's multiple comparison test was used. *** $p < 0.001$.

mutants of *mib1-2* and *mib1-3* (the *mib1-1* mutant has about two more primary branches than WT; Supplementary Figure 1). However, the secondary branches in the three *mib1* alleles increased significantly (Supplementary Figure 1). Thus, the increased branch number in the mutants was caused by accelerated bud outgrowth and sustained branching capacity among early developed primary branches. Additionally, the *mib1* mutants had a compact plant architecture, compared with WT (Figure 1E and Supplementary Figure 1A).

The flowers and young pods of the *mib1* mutants were smaller than those of WT (Supplementary Figure 2). The matured pods of the mutants were shorter, with decreased seed number and size (Figures 2A–E). The pod length of the *mib1* mutants (6.6± 0.03, 6.7± 0.05, and 7.3 ± 0.09 cm, respectively) was decreased, compared to that of WT (9.8 ± 0.11 cm). The seed number per pod of three *mib1* mutants (8.5 ± 0.11, 8.7 ± 0.20, and 9.9 ± 0.09, respectively) was much lower than that of WT (11.3 ± 0.65). Compared with the WT, mature seeds of *mib1* mutants were rounder and showed significantly decreased length, width, and thickness (Figures 2D, F). Therefore, the 100-seed weight was decreased by 26.82%, 18.63%, and 27.42% in *mib1-1*, *mib1-2*, and *mib1-3* mutants, compared with that of WT, respectively (Figure 2E).

The plant organ size is regulated by the coordination of two connected processes, cell division and expansion (Gonzalez et al., 2012). Microscopic examination of leaflet epidermal cells showed that the cell size decreased significantly in the *mib1-3* mutants in comparison with that of WT (Supplementary Figure 3A). The area of epidermal cells in the *mib1-3* mutants was only about half that in the WT plants (Supplementary Figure 3B), suggesting that *MIB1* augments organ size mainly by increasing the cell size.

Molecular characterization of the *MIB1* gene in mungbean

Genetic analysis of the *mib1* mutants was conducted by backcrossing *mib1-3* mutants with the WT plants. All F_1 plants were similar to WT. In the F_2 population, the WT plants and mutant plants segregated with a 3:1 ratio (87 WT plants and 25 mutants, $\chi^2 = 0.42 < \chi^2_{0.05} = 3.84$), indicating that *mib1* was a single recessive locus.

We conducted map-based cloning to identify the *MIB1* gene (Jiao et al., 2016). The *MIB1* gene was preliminarily mapped on chromosome 1 of the VC1973A genome (Kang et al., 2014),

linked with the markers ID244 and ID171 (Figure 3A). By developing new markers, the *mib1* mapping region was narrowed down to a 1.71-Mb region flanked by the markers ID218 and ID201 (Figure 3A). Based on the functional annotation (Kang et al., 2014) and the mutant phenotype, *Vradi01g10280* (*LOC106766026*) in the mapping region was identified as the candidate (Figure 3B). Sequencing of the PCR products of *Vradi01g10280* from WT and *mib1* mutants displayed mutations (Figure 3B), showing that three alleles, *mib1-1*, *mib1-2*, and *mib1-3*, carried different deletions (1-bp deletion, 1-bp deletion, and 21-bp deletion, respectively). qRT-PCR analysis of shoot buds (2 weeks after germination) revealed that there were decreased expression of the *Vradi01g10280* gene in the *mib1* mutants (Figure 3C).

Segregation analysis showed that 150 mutated plants out of a total of 642 individuals from the F_2 mapping population were homozygous for the 21-bp deletion in *Vradi02g10020*, indicating that the deletion co-segregates with the mutant phenotype. Therefore, *MIB1* (*Vradi01g10280*) encoded a member of MATE/DTX proteins (Figure 4), which was closely related to DTX54/BIGE1A in *Arabidopsis* and BIGE in maize (Suzuki et al., 2015), affecting plant architecture in mungbean.

The relative expression of the *MIB1* gene in different tissues of mungbean was analyzed by qRT-PCR. It was found that the *MIB1* gene was expressed in all tissues, with the highest expression level in the young inflorescence (Figure 3D).

MIB1 was a member of the MATE/DTX family proteins in mungbean

Multiple amino acid sequence alignments of the MIB1 protein (XP_014506278.1) with its homologs indicated that it shared a conserved MatE domain (Figure 4A). The MIB1 protein was predicted to have 12 transmembrane domains with N-termini towards the inside of the cell (Supplementary Figure 4).

We conducted a BLASTP search for sequences with homology to MIB1 to characterize the MATE/DTX family in the mungbean database (Kang et al., 2014) and found 56 MATE/DTX proteins in the mungbean genome (Supplementary Table 2). These mungbean MATE/DTX proteins were classified into four groups by phylogenetic analysis with *Arabidopsis* MATE/DTX proteins (Figure 4B; Wang et al., 2016). It was found that MIB1 had two other closely related homologs in mungbean, XP_022635913.1 (Vradi05g00900) and XP_014506743.1 (Vradi07g25110, Figure 4B).

In order to investigate the origin of MIB1 in legume plants, we identified MIB1 closed homologs from a number of public databases (Supplementary Table 2). The phylogenetic tree of aligned legume DTX54 and DTX55 orthologs was constructed (Figure 4B). It was found that one copy encoding the ortholog to DTX54 in legume formed the LegDTX54 clade, which was distinct from the LegDTX55 clade (Figure 4C). In contrast, within the LegDTX55 clade, there were different copies in

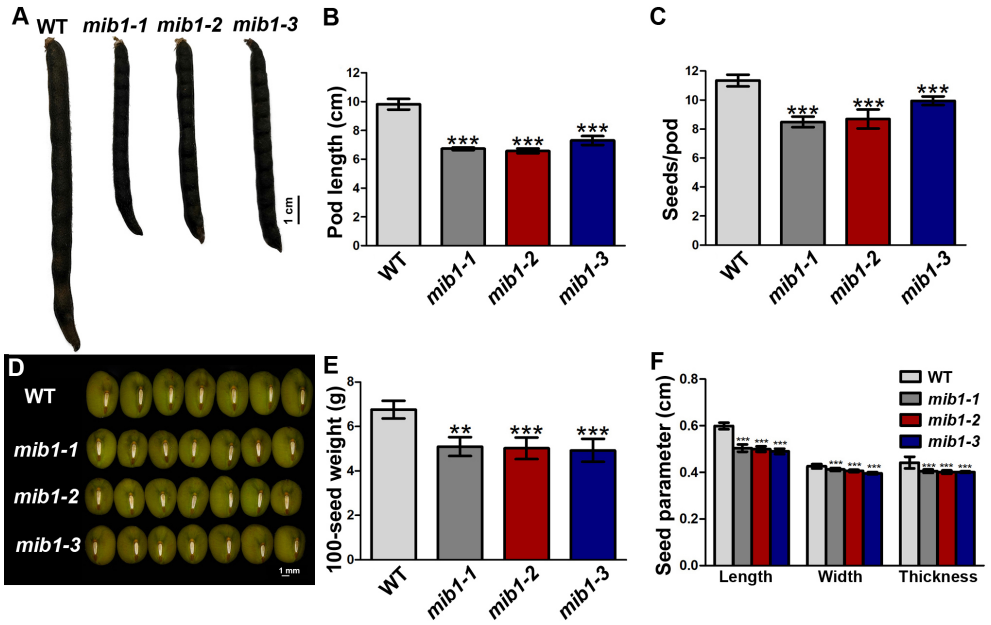


FIGURE 2
Characterization of pods and seeds of WT and *mib1* mutants. **(A)** Pods of WT and *mib1* mutants at matured stage. **(B)** Pod length of WT and *mib1* mutants (n = 150). **(C)** Seed number per pod of WT and *mib1* mutants (n = 150). **(D)** Seeds of WT and *mib1* mutants. **(E)** The 100-seed weights of WT and *mib1* mutants (n = 5). **(F)** Seed parameters of WT and *mib1* mutants (n = 200). The data were means ± SD. The Tukey's multiple comparison test was used. ** $p < 0.01$, *** $p < 0.001$.

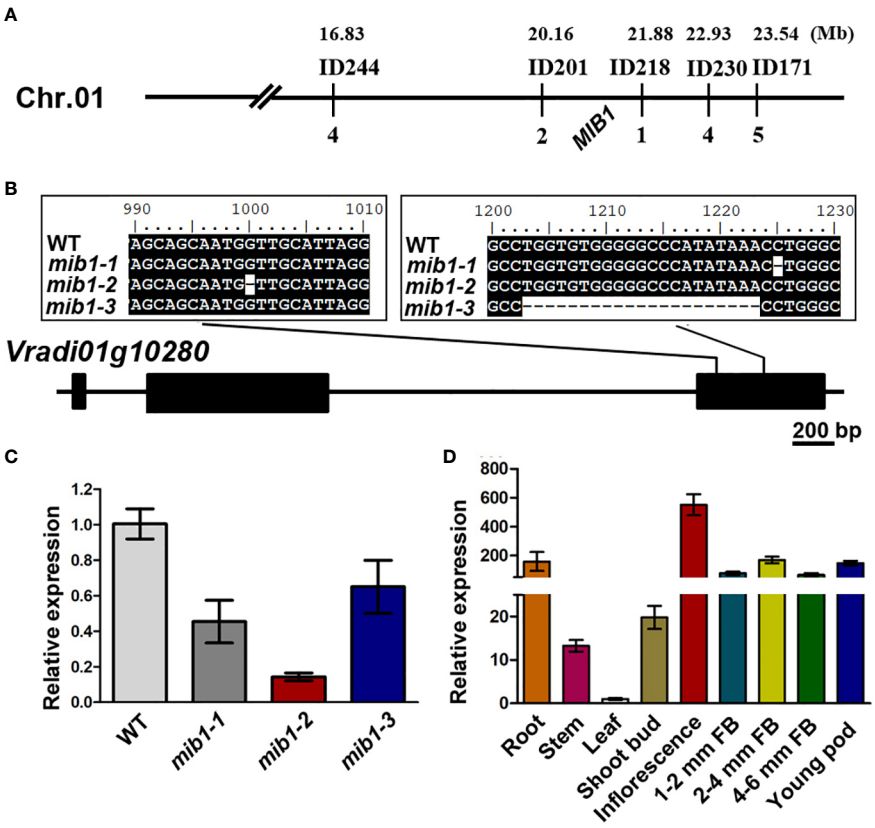


FIGURE 3
Map-based cloning of *MIB1*. **(A)** Genetic map of *MIB1* in mungbean. **(B)** Mutations in the open reading frame of *Vradi01g10280*. Numbers up the sequence indicate the position on the open reading frame. **(C)** Analysis of *MIB1* expression in shoot buds of WT and *mib1* mutant by qRT-PCR. **(D)** Relative expression level of *MIB1* in different tissues of WT.

legume, such as two copies in adzuki bean (*V. angularis*) and mungbean, and one copy in *Medicago truncatula* and *Lotus japonicus* (Figure 4C). The best phylogeny places the legume DTX55A (LegDTX55A) subclade and the legume DTX55B (LegDTX55B) subclade sister together, forming the LegDTX55 clade in legume (Figure 4C).

Heterologous expression of mungbean *MIB1* gene increased TBA tolerance in the mutant *Escherichia coli*

To investigate the functional character of the MIB1 protein, the expression vector containing the *MIB1* gene or empty vector was transformed into the WT strain (K12) and mutant strain (*ΔacrB)* in *E. coli*. The *ΔacrB* mutant strain lacks the functional AcrAB complex, the multidrug efflux carrier (Seo et al., 2012), and cannot grow under toxic conditions. The transformed cells were grown on the medium with and without different concentrations of TBA. The *ΔacrB* mutant cells with empty expressing vector (pET32a) did not grow on an LB plate with 10 and 15 mg/ml TBA (Figure 5A), while the *MIB1*-expressing *ΔacrB* cells continued their growth on the LB medium with 10 and 15 mg/ml TBA (Figure 5A), suggesting that MIB1 in mungbean potentially acts as a MATE/DTX family transporter.

In order to further verify the results of the plate experiment, we determined the growth curve of the strains under 0, 10, and 15 mg/ml TBA treatment in liquid LB medium (Figures 5B–D). Compared to those of expressing *MIB1* cells and the WT strain, TBA treatment significantly inhibited the growth of the mutant strain (Figures 5B–D). Under 10 and 15 mg/ml TBA treatments for 24 h, the growth curve of the mutant strain expressing *MIB1* was similar to those of the WT strain with and without expressing *MIB1* (Figures 5C, D). The above results showed that heterologous expression of mungbean *MIB1* increased TBA tolerance of the *ΔAcrB* mutant strain.

Heterologous expression of mungbean *MIB1* gene partially rescued the pod phenotype of *dtx54* mutant in *Arabidopsis*

It has been reported that loss of function of *DTX54/BIGE1A*, *MIB1* ortholog in *Arabidopsis*, leads to the shortened pods (Suzuki et al., 2015). To test whether *MIB1* performs a similar function to *DTX54/BIGE1A* in the control of pod development, the coding sequence of *MIB1* driven by the cauliflower mosaic virus 35S (CaMV35S) promoter was transformed into the *Arabidopsis dtx54* mutant (*35S::MIB1/dtx54*, two representative lines L04 and L06). It was found that the shortened pods were partially rescued in *35S::MIB1/dtx54* transgenic lines (Figures 6A, B). The expression of *MIB1* was detected in *35S::MIB1/dtx54* transgenic lines (Figure 6C). The above results indicated that the mungbean *MIB1* gene plays a conserved role in the pod development.

RNA-seq analysis of the wild-type plants and *mib1* mutants

To investigate the potential genes whose expression was altered in the mutants, RNA-sequencing (RNA-seq) analysis was conducted to compare the gene expression levels in shoot buds and young pods between WT and *mib1-3* mutants. A total of 3,173 and 875 differentially expressed genes (DEGs) were identified at the two developmental stages, respectively

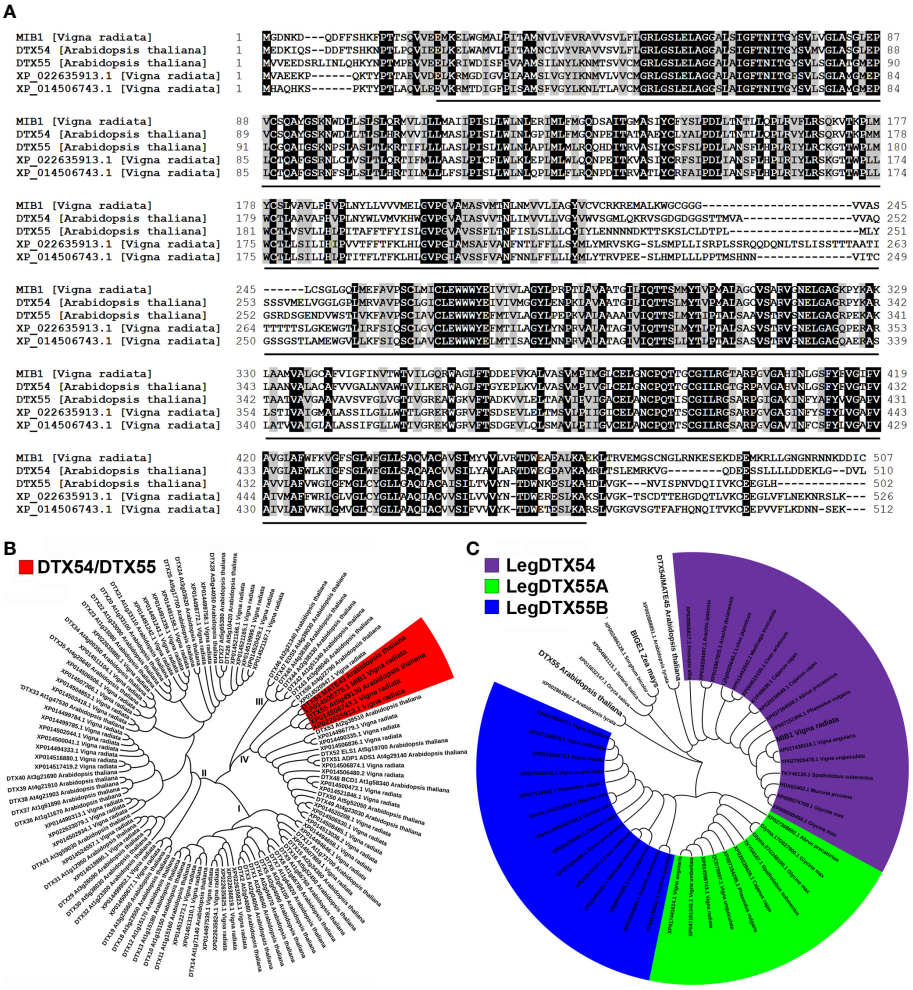


FIGURE 4
Analysis of the MIB1 protein isolated from mungbean. **(A)** Alignment of amino acid sequences of MIB1, DTX54, DTX55, XP_022635913.1, and XP_014506743.1. The conserved MatE domain was indicated by the black line. **(B)** Phylogenetic tree of MATE family proteins in mungbean and *Arabidopsis*. I, II, III, and IV represent four different groups. **(C)** Phylogenetic analysis of LegDTX54 clade and LegDTX55 clade proteins in legume.

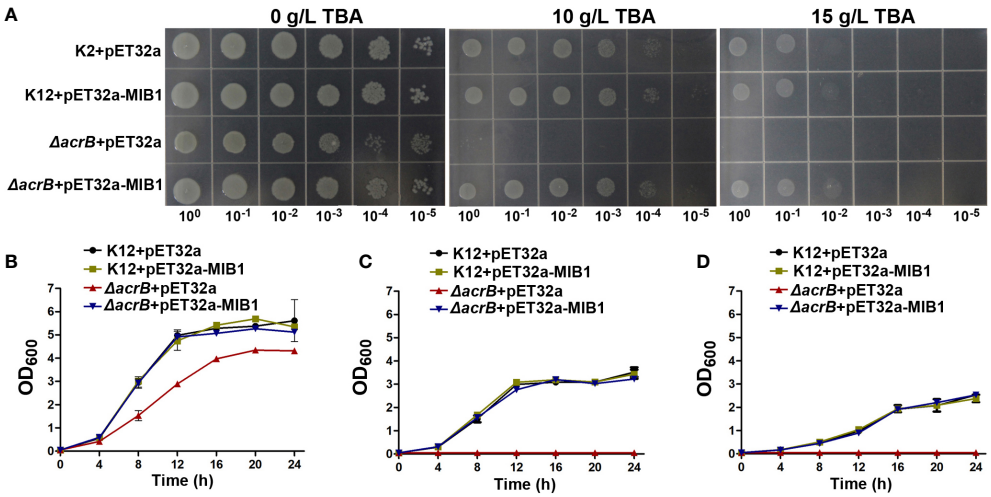


FIGURE 5
Analysis of MIB1 transport function in *Escherichia coli*. **(A)** The effect of *MIB1* expression on the growth of (*E*) *coli* cells under TBA treatment on the LB plate. (*E*) *coli* cells were spotted on the LB plate with 0, 10, and 15 g/L TBA for 24 h. 10^0, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, and 10^{-5} represented dilution series. **(B–D)** The effect of *MIB1* expression on the growth curve of (*E*) *coli* cells under TBA treatment. *(E) coli* cells were inoculated in liquid LB medium with 0 **(B)**, 10 **(C)**, and 15 **(D)** g/L TBA for 24 h. The data were means ± SD (n = 3).

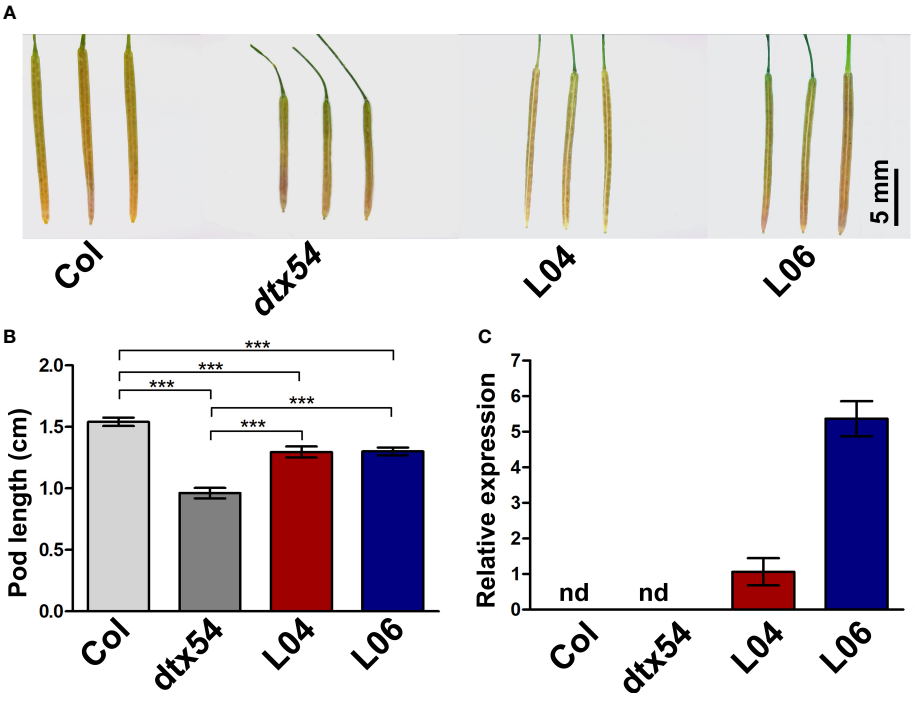


FIGURE 6
MIB1 partially rescued the shortened pod phenotype of *Arabidopsis dtx54* mutant. **(A)** The pod phenotype of the wild-type plant (Col), *dtx54* mutant, and *35S::MIB1* transgenic lines of *dtx54* (L04 and L06). **(B)** The pod length of Col, *dtx54* mutants, and two transgenic lines (n = 100). **(C)** qRT-PCR analysis of *MIB1* expression from Col, *dtx54* mutant, and two transgenic lines. nd, not detected. The data were means ± SD. One-way ANOVA was used. ***$p < 0.001$.

(Figure 7, Supplementary Tables 3 and 4). The qRT-PCR analysis confirmed the results of the RNA-Seq (Figures 8A, B).

Consistent with the mutant phenotype of plant architecture, the expression levels of key components of the plant hormone pathway and transcription factors related to plant development and growth were changed in the *mib1* mutants (Figures 8C, D). Among DEGs of the shoot buds between WT and *mib1* mutants, three auxin biosynthesis genes (*LOC106761734*, *LOC106761746*, and *LOC106764471*, *YUCCA 4*, *YUC4*) and a gene encoding auxin transporter (*LOC106760087*, *PIN-FORMED 1C*, *PIN1C*) were downregulated (Figure 8C). Several transcription factor encoding genes involved in plant development, such as *LOC06756958* (*Auxin Response Factor 2*, *ARF2*), *LOC106778614* (TCP family gene, *TCP4*), *LOC106769314* and *LOC106758337* (SPL family genes, *SPL7* and *SPL8*), and *LOC06765209* (AP2/ERF family gene *AINTEGUMENTA*, *ANT*), were downregulated (Figure 8C). Additionally, *LOC106767323* (*DWARF 14*, *D14*), encoding a key component of the SL signaling pathway (Zhou et al., 2013), was downregulated in the shoot buds of the *mib1* mutants (Figure 8C).

It has been shown that auxin and cytokinin pathways play a key role in the control of pod development and seed number per pod (Liu et al., 2021; Qadir et al., 2021; Yu et al., 2022). We found that the auxin biosynthesis gene (*LOC106778822*, *YUC11*) and the auxin transporter encoding gene (*LOC106761294*, *PIN2*) were downregulated in the young pods of the *mib1* mutants (Figure 8D). Consistently, there was a significant reduction in IAA level in young pods of the *mib1* mutants, compared to that of WT (Supplementary Figure 5). In addition, the expression levels of *LOC106759647* (*Cytokinin dehydrogenase 3*, *CKX3*) and *LOC106764037* (*LONELY GUY 3*, *LOG3*), related to the cytokinin pathway, were also significantly changed in the young pods of the *mib1* mutants (Figure 8D).

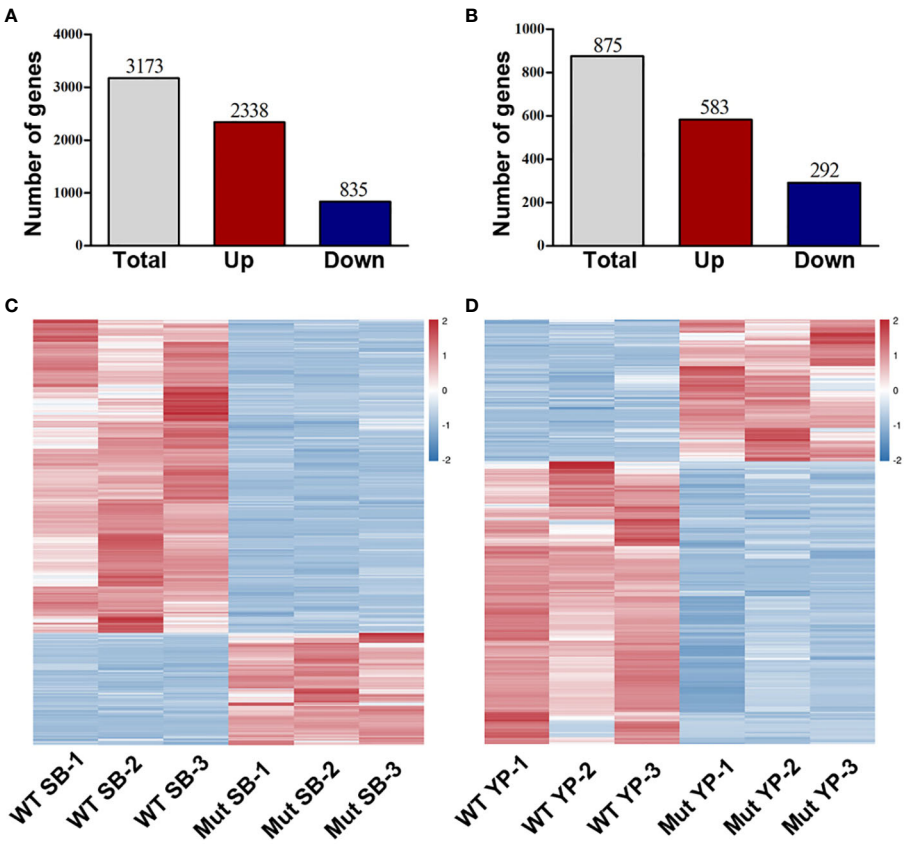


FIGURE 7
RNA-seq analysis of DEGs between WT and *mib1* mutants. **(A)** Number of DEGs of the shoot buds between WT and *mib1* mutants. **(B)** Number of DEGs of the young pods between WT and *mib1* mutants. **(C)** Heat map of the DEGs of the shoot buds between WT and *mib1* mutants. **(D)** Heat map of the DEGs of the young pods between WT and *mib1* mutants.

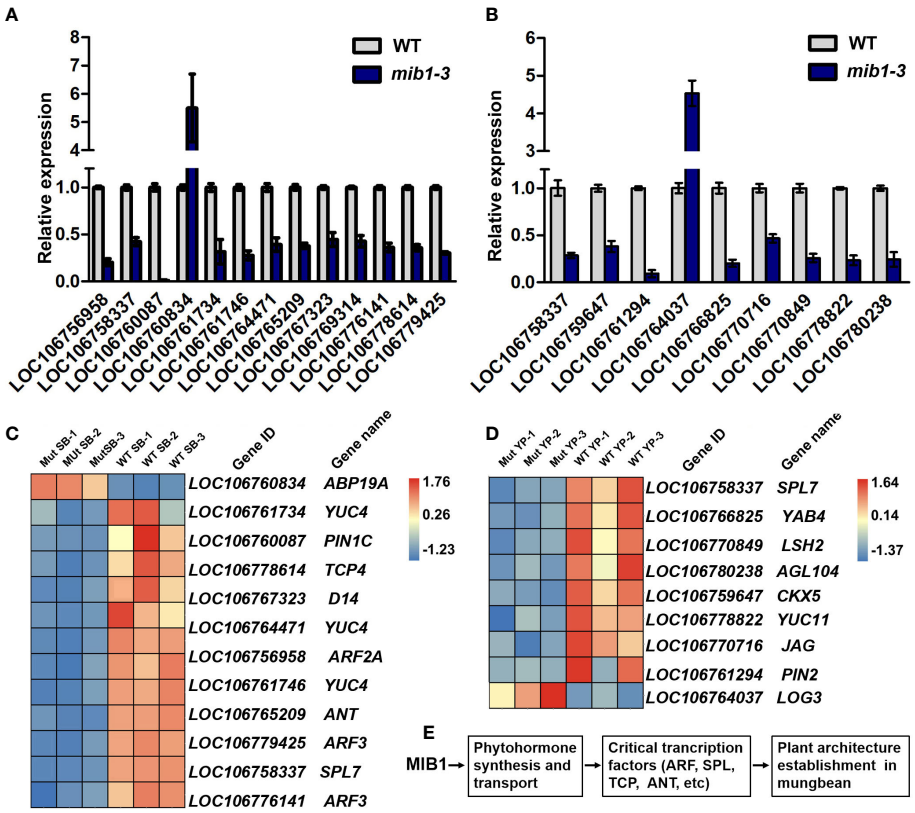


FIGURE 8
Critical DEGs involved in the control of plant architecture in mungbean. **(A)** qRT-PCR analysis of 13 DEGs in the shoot buds between WT and *mib1* mutants. **(B)** qRT-PCR analysis of eight DEGs in young pods between WT and *mib1* mutants. **(C)** Heat map showing critical DEGs involved in the control of the rate of leaf production, branch number, and organ size. **(D)** Heat map showing critical DEGs in involved young pod development. **(E)** A hypothetical model of MIB1 affecting plant architecture in mungbean.

Discussion

MIB1 encoded a MATE/DTX family transporter, affecting plant architecture in mungbean

The plant architecture significantly affects the seed yield of grain legume. However, the underlying molecular mechanism is still poorly understood (Liu et al., 2020). In this study, The mutations of the *MIB1* gene in mungbean resulted in bushy and compact plant architecture (Figure 1) and shortened pods with smaller and rounder seeds (Figure 2). Map-based cloning showed that the *MIB1* gene encoded a MATE/DTX family protein in mungbean, which was an ortholog of DTX54/BIGE1A in *Arabidopsis* and BIGE in maize (Figure 4). It has been reported that loss of function of the *DTX54/BIGA1A* gene, *MIB1* ortholog in *Arabidopsis*, gives rise to increased branch numbers and shortened pods (Suzuki et al., 2015). We found that heterologous expression of the *MIB1* gene partially rescued the phenotype of *dtx54/bige1a* mutant in *Arabidopsis*, suggesting that MIB1 plays a conserved role in the control of pod development.

MIB1 belonged to group IV of the MATE/DTX family (Figure 4A). Complementation assays in *E. coli* showed that MIB1 potentially acted as a MATE/DTX transporter in mungbean. Meanwhile, there was a significant reduction in IAA levels in young pods of the *mib1* mutants (Supplementary Figure 5). Consistently, transcriptome analysis revealed that expression levels of the genes related to auxin synthesis and transport were decreased (Figure 8). Thus, our results suggested that auxin plays a key role in regulating plant architecture in mungbean. The alteration of plant architecture in the *mib1* mutants was probably due to the modulated levels of auxin and other plant hormones and then the altered expression of the downstream genes related to plant growth and development (Figure 8E).

It has been reported that the group IV MATE/DTX transporters are able to modulate plant hormone levels such as auxin and ABA in *Arabidopsis* and rice (Li et al., 2014; Zhang et al., 2014; Qin et al., 2021). Thus, how plant hormone level is

modulated by the MIB1 protein should be investigated in more detail in the future.

Phylogenetic analysis of DTX54 and DTX55 orthologs in legume

It has been shown that the *DTX54/BIGE1A* and *DTX55/BIGE1B* in *Arabidopsis*, two paralogs, have partial functional redundancy and diversity (Suzuki et al., 2015). The mutant of the *Arabidopsis DTX54/BIGE1A* gene shows increased number of leaves (Suzuki et al., 2015). By contrast, the *dtx55* mutant exhibits a slight increase in leaf number, suggesting that *Arabidopsis DTX54* has a greater role in the control of leaf initiation, while the leaf number of the double mutants of *DTX54/BIGE1A* and *DTX55/BIGE1B* is enhanced compared to the single mutants, indicating that there is an additive interaction between the two genes (Suzuki et al., 2015).

It was found that there were 56 MATE/DTX family proteins in mungbean genome (Figure 4), among which two other MATE/DTX proteins are closely related to MIB1 and might redundantly affect plant development and growth in mungbean. Moreover, based on the public sequences, we identified DTX54 and DTX55 orthologs in legume. We found that there was a single copy encoding the DTX54 orthologs in legume (Figure 4). In contrast, the legDTX55 clade in legume could be further divided into two subclades, LegDTX55A and LegDTX55B (Figure 4B).

At present, mutant libraries for several legume species such as *M. truncatula*, *L. japonicus*, and *G. max* are available (Tadege et al., 2008; Małolepszy et al., 2016; Li et al., 2017), and it would be worth identifying the mutant lines of the LegDTX54 and LegDTX55 clade genes in these species to dissect their function in plant architecture establishment. Moreover, it is also interesting to study the interactions between the LegDTX54 clade and LegDTX55 clade genes during plant development and growth in legume.

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

XL wrote the manuscript. XL, YJ, MS, ZJ, HZ, DQ, QC, YX, and XY performed the experiments. XC and ZS supervised the research. XL and ZS analyzed the data and prepared the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This research was funded by the Science Foundation of Jiangsu Province, China (BE2021718), Jiangsu Seed Industry Revitalization Project (JBGS[2021]004), the Jiangsu Agricultural Science and Technology Innovation Fund of China (CX(20)3030), and the Students' innovation and entrepreneurship training program of National experimental Teaching Center for Plant Production (ZKF202212).

Acknowledgments

We would like to thank Professor Chuanzhen Jiang (South China Agricultural University) for providing the K12 and *ΔacrB* mutant strains.

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.1064685/full#supplementary-material

SUPPLEMENTARY FIGURE 1
The number of primary and secondary branches of WT and *mib1* mutants at the matured stage. **(A)** Plant architecture of WT and *mib1* mutants at the 8 weeks after germination; **(B)** The number of primary and secondary branches of WT and *mib1* mutants. The data were means ± SD (n=10). The Tukey's multiple comparison test was used. $^{**}p < 0.01$, $^{***}p < 0.001$.

SUPPLEMENTARY FIGURE 2
Characterization of the flowers **(A)** and young pods **(B)** of WT and *mib1* mutants.

SUPPLEMENTARY FIGURE 3
SEM analysis of the terminal leaflets of WT and *mib1* mutants. **(A)** The epidermal cells of the terminal leaflets of the fifth compound leaves in WT and *mib1-3* mutants; **(B)** The size of epidermal cells from the terminal

leaflets of WT and *mib1-3* mutants. The data were means ± SD (n=100). The Student's test was used. ***$p < 0.001$.

SUPPLEMENTARY FIGURE 4
The 12 transmembrane domains of the MIB1 protein was predicted using the DeepTMHMM. **(A)** Most likely topology of MIB1; **(B)** Posterior probabilities of MIB1.

SUPPLEMENTARY FIGURE 5
The hormone levels of IAA and ABA in young pods of WT and *mib1* mutants. The data were means ± SD (n=5). The Student's test was used. **$p < 0.01$.

SUPPLEMENTARY TABLE 1
Primer list used in this study.

SUPPLEMENTARY TABLE 2
MATE family proteins used for phylogenetic analysis in this study.

SUPPLEMENTARY TABLE 3
DEGs in shoot buds between WT and *mib1* mutants.

SUPPLEMENTARY TABLE 4
DEGs in young pods between WT and *mib1* mutants.

References

Audic, S., and Claverie, J. M. (1997). The significance of digital gene expression profiles. *Genome Res.* 7, 986–995. doi: 10.1101/gr.7.10.986

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

Burko, Y., Geva, Y., Refael-Cohen, A., Shleizer-Burko, S., Shani, E., Berger, Y., et al. (2011). From organelle to organ: ZRIZI MATE-type transporter is an organelle transporter that enhances organ initiation. *Plant Cell Physiol.* 52, 518–527. doi: 10.1093/pcp/pcr007

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

Clough, S. J., and Bent, A. F. (1998). Floral dip: A simplified method for agrobacterium-mediated transformation of *Arabidopsis thaliana*. *Plant J.* 16, 735–743. doi: 10.1046/j.1365-313x.1998.00343.x

Diener, A. C., Fink, G. R., and Gaxiola, R. A. (2001). Arabidopsis ALF5, a multidrug efflux transporter gene family member, confers resistance to toxins. *Plant Cell* 13, 1625–1637. doi: 10.1105/tpc.010035

Duan, W., Lu, F., Cui, Y., Zhang, J., Du, X., Hu, Y., et al. (2022). Genome-wide identification and characterisation of wheat MATE genes reveals their roles in aluminium tolerance. *Int. J. Mol. Sci.* 23, 4418. doi: 10.3390/ijms23084418

Gani, U., Nautiyal, A. K., Kundan, M., Rout, B., Pandey, A., and Misra, P. (2022). Two homeologous MATE transporter genes, *NtMATE21* and *NtMATE22*, are involved in the modulation of plant growth and flavonol transport in *Nicotiana tabacum*. *J. Exp. Bot.* 73 (18), 6186–6206. doi: 10.1093/jxb/erac249

Gao, J., Yang, S., Cheng, W., Fu, Y., Leng, J., Yuan, X., et al. (2017). *GmILPA1*, encoding an APC8-like protein, controls leaf petiole angle in soybean. *Plant Physiol.* 174, 1167–1176. doi: 10.1104/pp.16.00074

Gonzalez, N., Vanhaeren, H., and Inzé, D. (2012). Leaf size control: complex coordination of cell division and expansion. *Trends Plant Sci.* 17, 332–340. doi: 10.1016/j.tplants.2012.02.003

Graham, P. H., and Vance, C. P. (2003). Legumes: importance and constraints to greater use. *Plant Physiol.* 131, 872–877. doi: 10.1104/pp.017004

Guo, W., Chen, L., Herrera-Estrella, L., Cao, D., and Tran, L. P. (2020). Altering plant architecture to improve performance and resistance. *Trends Plant Sci.* 25, 1154–1170. doi: 10.1016/j.tplants.2020.05.009

Ishihara, T., Sekine, K. T., Hase, S., Kanayama, Y., Seo, S., Ohashi, Y., et al. (2008). Overexpression of the *Arabidopsis thaliana EDS5* gene enhances resistance to viruses. *Plant Biol. (Stuttg)* 10, 451–461. doi: 10.1111/j.1438-8677.2008.00050.x

Jia, M., Liu, X., Xue, H., Wu, Y., Shi, L., Wang, R., et al. (2019). Noncanonical ATG8-ABS3 interaction controls senescence in plants. *Nat. Plants* 5, 212–224. doi: 10.1038/s41477-018-0348-x

Jiao, K. Y., Li, X., Guo, W. X., Yuan, X. X., Cui, X. Y., and Chen, X. (2016). Genome re-sequencing of two accessions and fine mapping the locus of *lobed leaflet margins* in mungbean. *Mol. Breed* 36, 128. doi: 10.1007/s11032-016-0552-1

Jiao, K., Li, X., Su, S., Guo, W., Guo, Y., Guan, Y., et al. (2019). Genetic control of compound leaf development in mungbean (*Vigna radiata* l.). *Hortic. Res.* 6, 23. doi: 10.1038/s41438-018-0088-0

Kang, Y. J., Kim, S. K., Kim, M. Y., Lestari, P., Kim, K. H., Ha, B. K., et al. (2014). Genome sequence of mungbean and insights into evolution within *Vigna* species. *Nat. Commun.* 5, 5443. doi: 10.1038/ncomms6443

Kerr, S. C., Patil, S. B., de Saint Germain, A., Pillot, J. P., Saffar, J., Ligerot, Y., et al. (2021). Integration of the SMXL/D53 strigolactone signalling repressors in the model of shoot branching regulation in *Pisum sativum*. *Plant J.* 107, 1756–1770. doi: 10.1111/tpj.15415

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

Letunic, I., and Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* 44, W242–W245. doi: 10.1093/nar/gkw290

Li, L., He, Z., Pandey, G. K., Tsuchiya, T., and Luan, S. (2002). Functional cloning and characterization of a plant efflux carrier for multidrug and heavy metal detoxification. *J. Biol. Chem.* 277, 5360–5368. doi: 10.1074/jbc.M108777200

Li, Z., Jiang, L., Ma, Y., Wei, Z., Hong, H., Liu, Z., et al. (2017). Development and utilization of a new chemically-induced soybean library with a high mutation density. *J. Integr. Plant Biol.* 59, 60–74. doi: 10.1111/jipb.12505

Li, R., Li, J., Li, S., Qin, G., Novák, O., Pěnčík, A., et al. (2014). ADP1 affects plant architecture by regulating local auxin biosynthesis. *PloS Genet.* 10, e1003954. doi: 10.1371/journal.pgen.1003954

Li, X., Sun, M., Jia, Y., Qiu, D., Peng, Q., and Zhuang, L. (2022). Genetic control of the lateral petal shape and identity of asymmetric flowers in mungbean (*Vigna radiata* l.). *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.996239

Liu, L. M., Zhang, H. Q., Cheng, K., and Zhang, Y. M. (2021). Integrated bioinformatics analyses of PIN1, CKX, and yield-related genes reveals the molecular mechanisms for the difference of seed number per pod between soybean and cowpea. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.749902

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

Lu, P., Magwanga, R. O., Kirungu, J. N., Hu, Y., Dong, Q., Cai, X., et al. (2019). Overexpression of cotton a DTX/MATE gene enhances drought, salt, and cold stress tolerance in transgenic *Arabidopsis*. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.00299

Magalhaes, J. V., Liu, J., Guimarães, C. T., Lana, U. G., Alves, V. M., Wang, Y. H., et al. (2007). A gene in the multidrug and toxic compound extrusion (MATE) family confers aluminum tolerance in sorghum. *Nat. Genet.* 39, 1156–1161. doi: 10.1038/ng2074

Małolepszy, A., Mun, T., Sandal, N., Gupta, V., Dubin, M., Urbański, D., et al. (2016). The LORE1 insertion mutant resource. *Plant J.* 88, 306–317. doi: 10.1111/tpj.13243

Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-seq. *Nat. Methods* 5, 621–628. doi: 10.1038/nmeth.1226

Nawrath, C., Heck, S., Parinthawong, N., and Métraux, J. P. (2002). EDS5, an essential component of salicylic acid-dependent signaling for disease resistance in *Arabidopsis*, is a member of the MATE transporter family. *Plant Cell.* 14, 275–286. doi: 10.1105/tpc.010376

Nimmy, M. S., Kumar, V., Suthanthiram, B., Subbaraya, U., Nagar, R., Bharadwaj, C., et al. (2022). A systematic phylogenomic classification of the multidrug and toxic compound extrusion transporter gene family in plants. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.774885

Pan, X., Welti, R., and Wang, X. (2010). Quantitative analysis of major plant hormones in crude plant extracts by high-performance liquid chromatography-mass spectrometry. *Nat. Protoc.* 5, 986–992. doi: 10.1038/nprot.2010.37

Qadir, M., Wang, X., Shah, S., Zhou, X. R., Shi, J., and Wang, H. (2021). Molecular network for regulation of ovule number in plants. *Int. J. Mol. Sci.* 22, 12965. doi: 10.3390/ijms222312965

Qin, P., Zhang, G., Hu, B., Wu, J., Chen, W., Ren, Z., et al. (2021). Leaf-derived ABA regulates rice seed development *via* a transporter-mediated and temperature-sensitive mechanism. *Sci. Adv.* 7, eabc8873. doi: 10.1126/sciadv.abc8873

Rameau, C., Bertheloot, J., Leduc, N., Andrieu, B., Foucher, F., and Sakr, S. (2015). Multiple pathways regulate shoot branching. *Front. Plant Sci.* 5. doi: 10.3389/fpls.2014.00741

Reinhardt, D., and Kuhlemeier, C. (2002). Plant architecture. *EMBO Rep.* 3, 846–851. doi: 10.1093/embo-reports/kvf177

Rogers, E. E., and Guerinot, M. L. (2002). FRD3, a member of the multidrug and toxin efflux family, controls iron deficiency responses in *Arabidopsis*. *Plant Cell* 14, 1787–1799. doi: 10.1105/tpc.001495

Seo, P. J., Park, J., Park, M. J., Kim, Y. S., Kim, S. G., Jung, J. H., et al. (2012). A golgi-localized MATE transporter mediates iron homoeostasis under osmotic stress in *Arabidopsis*. *Biochem. J.* 442, 551–561. doi: 10.1042/BJ20111311

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

Suzuki, M., Sato, Y., Wu, S., Kang, B. H., and McCarty, D. R. (2015). Conserved functions of the MATE transporter BIG EMBRYO1 in regulation of lateral organ size and initiation rate. *Plant Cell.* 27, 2288–2300. doi: 10.1105/tpc.15.00290

Tadege, M., Wen, J., He, J., Tu, H., Kwak, Y., Eschstruth, A., et al. (2008). Large-Scale insertional mutagenesis using the Tnt1 retrotransposon in the model legume *Medicago truncatula*. *Plant J.* 54, 335–347. doi: 10.1111/j.1365-313X.2008.03418.x

Thompson, E. P., Wilkins, C., Demidchik, V., Davies, J. M., and Glover, B. J. (2010). An arabidopsis flavonoid transporter is required for anther dehiscence and pollen development. *J. Exp. Bot.* 61, 439–451. doi: 10.1093/jxb/erp312

Upadhyay, N., Kar, D., and Datta, S. A. (2020). Multidrug and toxic compound extrusion (MATE) transporter modulates auxin levels in root to regulate root development and promotes aluminium tolerance. *Plant Cell Environ.* 43, 745–759. doi: 10.1111/pce.13658

Upadhyay, N., Kar, D., Deepak Mahajan, B., Nanda, S., Rahiman, R., Panchakshari, N., et al. (2019). The multitasking abilities of MATE transporters in plants. *J. Exp. Bot.* 70, 4643–4656. doi: 10.1093/jxb/erz246

Wang, L., Bei, X., Gao, J., Li, Y., Yan, Y., and Hu, Y. (2016). The similar and different evolutionary trends of MATE family occurred between rice and *Arabidopsis thaliana*. *BMC Plant Biol.* 16, 207. doi: 10.1186/s12870-016-0895-0

Wang, Y., and Li, J. (2008). Molecular basis of plant architecture. *Annu. Rev. Plant Biol.* 59, 253–279. doi: 10.1146/annurev.arplant.59.032607.092902

Yu, S. X., Jiang, Y. T., and Lin, W. H. (2022). Ovule initiation: the essential step controlling offspring number in *Arabidopsis*. *J. Integr. Plant Biol.* 64, 1469–1486. doi: 10.1111/jipb.13314

Zhang, H., Zhu, H., Pan, Y., Yu, Y., Luan, S., and Li, L. (2014). A DTX/MATE-type transporter facilitates abscisic acid efflux and modulates ABA sensitivity and drought tolerance in *Arabidopsis*. *Mol. Plant* 7, 1522–1532. doi: 10.1093/mp/ssu063

Zhou, F., Lin, Q., Zhu, L., Ren, Y., Zhou, K., Shabek, N., et al. (2013). D14-SCF (D3)-dependent degradation of D53 regulates strigolactone signalling. *Nature* 504, 406–410. doi: 10.1038/nature12878

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 11 January 2023
DOI 10.3389/fpls.2022.1094112

OPEN ACCESS

EDITED BY
Jianghua Chua,
Key Laboratory of Tropical Plant Resource and Sustainable Use (CAS), China

REVIEWED BY
Zhaoming Qi,
Northeast Agricultural University, China
Hengyou Zhang,
Northeast Institute of Geography and Agroecology (CAS), China

*CORRESPONDENCE
Yanbo Cheng
ybcheng@scau.edu.cn
Hai Nian
hnian@scau.edu.cn

†These authors share first authorship

SPECIALTY SECTION
This article was submitted to Functional and Applied Plant Genomics, a section of the journal Frontiers in Plant Science

RECEIVED 09 November 2022
ACCEPTED 15 December 2022
PUBLISHED 11 January 2023

CITATION
Luo S, Jia J, Liu R, Wei R, Guo Z, Cai Z, Chen B, Liang F, Xia Q, Nian H and Cheng Y (2023) Identification of major QTLs for soybean seed size and seed weight traits using a RIL population in different environments. *Front. Plant Sci.* 13:1094112. doi: 10.3389/fpls.2022.1094112

Identification of major QTLs for soybean seed size and seed weight traits using a RIL population in different environments

Shilin Luo[1,2,3†], Jia Jia[1,2,3†], Riqian Liu[1,2,3†], Ruqian Wei[1,2,3], Zhibin Guo[1,2,3], Zhandong Cai[1,2,3], Bo Chen[1,2,3], Fuwei Liang[1,2,3], Qiuju Xia[4], Hai Nian[1,2,3*] and Yanbo Cheng[1,2,3*]

[1]The State Key Laboratory for Conservation and Utilization of Subtropical Agro-bioresources, South China Agricultural University, Guangzhou, Guangdong, China, [2]The Key Laboratory of Plant Molecular Breeding of Guangdong Province, College of Agriculture, South China Agricultural University, Guangzhou, Guangdong, China, [3]Guangdong Laboratory for Lingnan Modern Agriculture, Guangzhou, Guangdong, China, [4]Rice Molecular Breeding Institute, Granlux Associated Grains, Shenzhen, Guangdong, China

Introduction: The seed weight of soybean [*Glycine max* (L.) Merr.] is one of the major traits that determine soybean yield and is closely related to seed size. However, the genetic basis of the synergistic regulation of traits related to soybean yield is unclear.

Methods: To understand the molecular genetic basis for the formation of soybean yield traits, the present study focused on QTLs mapping for seed size and weight traits in different environments and target genes mining.

Results: A total of 85 QTLs associated with seed size and weight traits were identified using a recombinant inbred line (RIL) population developed from Guizao1×B13 (GB13). We also detected 18 environmentally stable QTLs. Of these, *qSL-3-1* was a novel QTL with a stable main effect associated with seed length. It was detected in all environments, three of which explained more than 10% of phenotypic variance (PV), with a maximum of 15.91%. In addition, *qSW-20-3* was a novel QTL with a stable main effect associated with seed width, which was identified in four environments. And the amount of phenotypic variance explained (PVE) varied from 9.22 to 21.93%. Five QTL clusters associated with both seed size and seed weight were summarized by QTL cluster identification. Fifteen candidate genes that may be involved in regulating soybean seed size and weight were also screened based on gene function annotation and GO enrichment analysis.

Discussion: The results provide a biologically basic reference for understanding the formation of soybean seed size and weight traits.

KEYWORDS

soybean, seed size, seed weight, stable QTLs, QTL clusters

1 Introduction

Soybean is one of the globally esteemed food crops with numerous nutritionalsubstances including protein, carbohydrates, lipids, minerals, vitamins, and bioactive substances such as isoflavones, saponins, sterols and phospholipids. With its vital values in agriculture and economy, soybean plays an important role in agricultural and industrial activities such as processing food and feed and producing textiles. While the demand for soybean has been increasing globally, soybean yield enhancement is now receiving significant attention for its potential for evolving productivity.

Seed size and seed weight are quantitative traits controlled by multiple genes and constrained by environmental factors. The forming of both traits is a complex biological process (Li et al., 2020). Seed size usually measure s seed length, seed width, and seed thickness (Salas et al., 2006). Many studies have been done worldwide on QTL localization of seed size and weight in soybean. In contrast, studies on QTL localization for seed size were relatively less than that for seed weight. For example, on the soybean data website SoyBase (http://www.soybase.org/) as of March 2022, the number of QTL publications are 18 on seed length, 16 on seed width, and 23 on seed thickness versus the number of QTL publications on seed weight are 300.

Salas et al. (2006) detected a total of 19 QTLs associated with seed shape using three RIL populations in two environments. Xu et al. (2011) used MAJ Multi-QTL Joint Analysis (MAJ) and composite interval mapping (CIM) for QTLs mapping. A total of 121 main QTLs were detected in the $F_{2:3}$, $F_{2:4}$ and $F_{2:5}$ populations from the direct and reciprocal crosses of Lishuizhongzihuang with Nannong493-1. Sun et al. (2021) used a RIL population to detect QTLs for seed size traits in four environments. Ten QTLs controlling related traits were identified, of which five QTLs distributed on chromosomes 02, 04, 06, 13 and 16 were detected in at least two environments with PVE ranging from 3.6 to 9.4%. The results of similar studies (Salas et al., 2006; Xu et al., 2011; Hu et al., 2013; Niu et al., 2013; Yang et al., 2017; Teng et al., 2018; Gao et al., 2019; Cui et al., 2020; Sun et al., 2021; Kumawat and Xu, 2021) vary form different localization methods and different genetic backgrounds

The first study on QTL localization of seed weight in bean was reported by Sax (1923), in which he used seed colors as the markers. With the advancement of molecular marker technology, scholars had developed various molecular markers such as restriction fragment length polymorphism (RFLP), simple sequence repeat (SSR) and single nucleotide polymorphism (SNP) with different mapping populations to identify QTLs of soybean seed weight. In 1996, Maughan et al. (1996) detected 6 QTLs associated with seed weight in 150 F_2 populations and their 150 $F_{2:3}$ lines using 91 polymorphic genetic markers including RFP, RAPD and SSR. Zhang et al. (2016) performed a genome-wide association study (GWAS) using 309 soybean germplasm resources with 31045 SNP markers and associated QTLs related to seed weight to chromosome 04 and 19. More results from previous studies (Mian et al., 1996; Csanadi et al., 2001; Lee et al., 2001; Chapman et al., 2003; Fasoula et al., 2004; Hyten et al., 2004; D. Li et al., 2008; Han et al., 2012; Pathan et al., 2013; Kato et al., 2014; Yan et al., 2014; Kulkarni et al., 2017; Liu D. et al., 2018) are not identical, but these QTLs have laid a foundation for molecular marker-assisted breeding.

In summary, most researchers use different numbers of mapping populations to detect QTL in various environments. However, studies using more mapping populations had fewer experimental settings, in contrast to studies using fewer mapping populations tended to have more experimental environments. In this study, one mapping population is used to detect QTL in six environments suggesting a higher standard for the identification of stable QTL. In addition, the QTLs localization of soybean seed size and weight traits by RIL population in different environments and the mining of main effect genes can help us to understand the biological basis of soybean yield trait forming process, which has great theoretical and practical significance in guiding the molecular design breeding and in consistently enhancing soybean yield.

2 Materials and methods

2.1 Materials and planting

GB13 RIL population, comprising 248 lines, constructed with Guizao1 as the female parent and B13 as the male parent was planted on three replications in July 2018, July 2019 and July 2020 at the Teaching and Research Base of Zengcheng (23°14′N, 113°38′E), and in July 2019 and July 2020 at the Experimental Teaching Base of Guangzhou (23°10′N, 113°22′E) (designated as 18ZC, 19ZC, 20ZC, 19GZ and 20GZ, respectively). In addition, one F_1 and 248 $F_{8:11}$ RILs of the GB13 RIL population were developed using the single seed descent method (Jiang et al., 2019). The experimental materials were planted in a completely randomized zonal design with one row per strain, a row length of 2.0 m, a row spacing of 0.5 m, and a plant spacing of 0.1 m. Field management was carried out according to conventional standards, and no pest or disease outbreaks occurred during growth.

2.2 Phenotypic statistics and analysis

Seed length (SL), seed width (SW) and seed thickness (ST) of soybean seeds were measured using an electronic vernier caliper (Figure 1). Twenty seeds from each strain were randomly selected for seed size measurement, after which they were

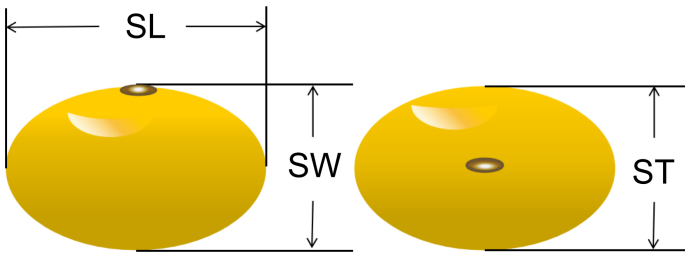


FIGURE 1
Measuring seed length (SL), seed width (SW) and seed thickness (ST).

weighed on an electronic balance. After three repetitions, five times the average of the weights were used as phenotypic data for the seed weight trait (100-seed weight, HSW).

The mean of the phenotypic data from the GB13 RIL population in five natural environments was calculated as the phenotypic data of the sixth environment (combined environment, designated as CE). Descriptive statistics, linear regression analysis and calculation of the skewness and the kurtosis were performed on phenotypic data from six environments using GraphPad Prism software (GraphPad Software Inc. v.7.0.4, United States), and the corresponding tables or figures were generated. Analysis of variance (ANOVA) and estimation of broad-sense heritability (h^2) were performed on phenotypic data from the five natural environments using Genstat 12_{th} Edition software. The formulae (Nyquist and Baker, 1991; Elattar et al., 2021) were as follows:

$$h^2 = \sigma_g^2/(\sigma_g^2 + \sigma_{ge}^2/n + \sigma_e^2/nr)$$

where σ^2_g denotes genotype variance, σ^2_{ge} denotes genotype × environment interaction variance, and σ^2_e denotes variance error. "n" denotes the number of environments, and "r" denotes the number of replicates per environment.

2.3 QTL localization for seed size and weight traits and analysis

The composite interval mapping method (Zeng, 1994) was used to map QTLs with a high-density bin linkage map previously constructed in the laboratory, as described, the map contained 56,561 SNPs and 3715 bin markers covering a total genome length of 3049.2 cM, with the length of the linkage cluster ranging from 120.22-211.37 cM, The average distance between markers was 0.80 cM, and the number of bin markers on each cluster varied from 147 to 259. (Jiang et al., 2019). The additive effect (ADD) and PVE of each QTL were analyzed by the WinQTLCart 2.5 software. The threshold of LOD for different traits was set to 2.5 at the 5% significance level, and LOD thresholds greater than 2.5 were considered as the basis for determining the presence of QTLs. QTLs located on the same chromosome with adjacent marker intervals were combined into one QTL if their confidence intervals overlapped. QTLs were named according to "q + trait + chromosome + sequence number", with the symbol "-" in between (Mccouch et al., 1997).

2.4 Stable QTLs and QTL clusters identification

Generally, QTLs detected in at least two environments were considered stable QTLs (Qi et al., 2017), but there were six experimental environments in this study, so QTLs detected in at least three environments were considered as stable QTLs in this study to identify the stable QTLs.

All QTLs were sorted by chromosome as the primary condition and physical position as the secondary condition. QTLs with identical chromosomes in overlapping or adjacent physical position (less than 5cM) were grouped into a cluster and identified as QTL cluster if it was associated with at least two traits (Liu et al., 2017; Wang et al., 2022). To mine the main effect and stable candidate genes, the identified QTL clusters contain at least one stable QTL in this study.

2.5 Candidate gene prediction

First, GO (Gene Ontology) term enrichment analysis was performed for the genes within the interval of the identified QTL clusters. This is useful for selecting candidate genes based on the sequence variation of genes between the parents. The GO annotation numbers were downloaded from the soybean data website SoyBase (https://www.soybase.org/), and then the WEGO2.0 online gene enrichment analysis mapping web page (http://wego.genomics.org.cn/) was used to create the gene enrichment cellular composition, molecular function and biological processes visualization diagrams. Next, genes that were highly or specifically expressed in soybean seeds were to be further screened. Gene expression data were obtained from the soybean data website SoyBase. Finally, the selected genes

were functionally annotated by the Phytozome database (https://phytozome-next.jgi.doe.gov/), and then the candidate genes were identified based on the relevant literature and related gene functions. Meanwhile, the heat map of candidate genes expression was employed to analyze tissue-specific expression (Jia et al., 2022) by the online site (https://www.omicshare.com/tools/Home/Soft/heatmap), and the structures of candidate genes were mapped using the online site of the Gene Structure Display Server (http://gsds.gao-lab.org/index.php). The resequencing data of the parental lines were referred to compare the variation of candidate genes between parental lines. The whole genome of parental lines were sequenced using the Illumina HiSeq X Ten p, with an average sequencing depth of 8× (Xian et al., 2022). High-quality sequencing data were assessed to predict the gene structural variations.

3 Result

3.1 Correlation analysis between seed size and seed weight

The regression linear analysis of seed size with seed weight shows that the SL, SW and ST were all highly significantly and positively correlated with HSW (p<0.01), with R-Squared (R^2) of regression linear analysis ranging from 0.440 to 0.804 (Figure 2). In 18ZC, 19ZC, 19GZ, 20ZC and CE, the R^2 between SW and HSW were the highest at 0.804, 0.721, 0.753, 0.552 and 0.742, respectively, while in 20GZ, the R^2 between ST and HSW was the highest at 0.754. Thus, seed size is indeed closely related to seed weight. 18ZC, 19ZC and CE had the next largest R^2 between SL and HSW at 0.685, 0.680 and 0.614, respectively and the smallest R^2 between ST and HSW at 0.679, 0.626 and 0.588, respectively. 19GZ and 20ZC had the next largest R^2 between ST and HSW at 0.713 and 0.467, and the smallest R^2 between SL and HSW at 0.664 and 0.440. 20GZ had the next largest R^2 between SL and HSW at 0.721, and the smallest R^2 between SW and HSW at 0.717. Also, the above results suggest that the correlation between seed size and seed weight seems to be applied in field selection breeding as a reference for selecting high-yielding varieties.

3.2 Descriptive statistics analysis of seed size and weight traits

The phenotypic data were analyzed by descriptive statistics, and the results showed that the SL, SW, ST (Supplementary Figure 1) and HSW of the female parent Guizao1 were higher than those of the male parent B13 with significant differences (Table 1), which laid the foundation for QTL mapping. The GB13 RIL population differed remarkably insize and weight traits, both of which showed transgressive segregation. The coefficient of variation (CV) for SL ranged from 4.33 to 5.25%, for SW from 3.96 to 4.78%, for ST from 4.71 to 5.18%, and for HSW from 10.68 to 12.98%. The CV for HSW was greater than that for SL, SW and ST, but they were relatively stable, which indicated that both parents contained genes that were acting in the phenotypic variation.

The absolute values of the skewness and the kurtosis of SL, SW, ST and HSW of the GB13 RIL population were less than 1 (Table 1). In addition, the frequency distribution graph (Figure 3) showed that the phenotypic data of seed size and weight traits displayed continuous variation. The above results indicated that seed size and weight traits of the GB13 RIL population obeyed normal distribution, which was consistent with the characteristics of the RIL population and belonged to quantitative genetic traits.

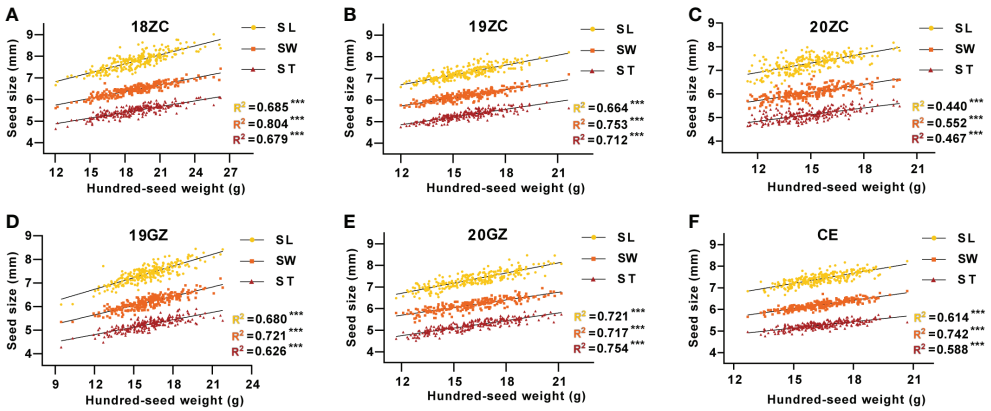


FIGURE 2
The correlation between seed size and seed weight of GB13 RIL population across the Teaching and Research Base of Zengcheng (ZC) in **(A)** 2018, **(B)** 2019 and **(C)** 2020, the Experimental Teaching Base of Guangzhou (GZ) in **(D)** 2019 and **(E)** 2020 and **(F)** their combined environment (CE). ***p < 0.001 (F-test). R^2 indicates the correlation coefficient.

10.3389/fpls.2022.1094112

TABLE 1 Statistical analysis ofseed size and weight traits for the parents and the lines at different environments.

Trait[#]	Env.[a]	Parents[b]		RILs[e]					
		Guizao1[c]	B13[d]	Range[f]	Mean[g]	SD[h]	CV (%)[i]	Skew.[j]	Kurt.[k]
SL	18ZC	8.34 ± 0.02	7.40 ± 0.01	6.66-9.01	7.77	0.41	5.25	0.12	-0.20
	19GZ	7.40 ± 0.02	6.90 ± 0.02	6.08-8.44	7.40	0.38	5.15	-0.19	0.66
	19ZC	7.40 ± 0.02	6.68 ± 0.02	6.52-8.21	7.27	0.32	4.33	0.32	-0.20
	20GZ	7.61 ± 0.01	6.86 ± 0.03	6.48-8.46	7.35	0.37	5.04	0.20	0.03
	20ZC	7.87 ± 0.03	6.85 ± 0.01	6.29-8.17	7.30	0.35	4.82	-0.35	-0.20
	CE	7.72 ± 0.07	6.94 ± 0.05	6.64-8.23	7.42	0.28	3.80	0.07	-0.03
SW	18ZC	6.65 ± 0.01	6.24 ± 0.01	5.59-7.43	6.46	0.29	4.44	-0.17	0.40
	19GZ	5.99 ± 0.02	5.84 ± 0.05	5.35-7.18	6.18	0.30	4.78	-0.01	0.22
	19ZC	6.02 ± 0.02	5.77 ± 0.02	5.65-7.18	6.20	0.25	3.96	0.31	0.46
	20GZ	5.97 ± 0.01	6.08 ± 0.02	5.37-6.95	6.18	0.28	4.55	-0.12	-0.23
	20ZC	6.10 ± 0.02	5.68 ± 0.03	5.37-6.75	6.05	0.27	4.50	-0.01	-0.27
	CE	6.34 ± 0.11	6.19 ± 0.08	5.66-6.85	6.21	0.21	3.39	0.14	-0.21
ST	18ZC	5.53 ± 0.01	5.22 ± 0.01	4.65-6.11	5.49	0.26	4.79	-0.22	0.07
	19GZ	5.17 ± 0.02	4.84 ± 0.04	4.28-5.90	5.24	0.25	4.83	-0.23	0.60
	19ZC	5.12 ± 0.01	4.90 ± 0.01	4.75-5.90	5.30	0.24	4.49	0.01	-0.47
	20GZ	5.11 ± 0.01	5.06 ± 0.01	4.52-5.88	5.23	0.27	5.18	-0.14	-0.39
	20ZC	5.22 ± 0.01	4.88 ± 0.02	4.61-5.81	5.12	0.24	4.71	0.25	-0.20
	CE	5.16 ± 0.01	5.05 ± 0.05	4.76-5.75	5.27	0.18	3.39	-0.10	-0.06
HSW	18ZC	18.85 ± 0.1	16.32 ± 0.06	12.05-26.22	18.88	2.45	12.98	0.14	0.28
	19GZ	15.35 ± 0.19	12.90 ± 0.24	9.45-21.80	16.09	1.91	11.89	0.02	0.61
	19ZC	14.97 ± 0.06	12.32 ± 0.05	12.00-21.63	15.77	1.68	10.68	0.29	0.02
	20GZ	15.40 ± 0.05	13.27 ± 0.13	11.70-21.17	16.08	2.01	12.52	0.14	-0.65
	20ZC	17.30 ± 0.12	13.02 ± 0.09	11.43-20.03	14.96	1.75	11.71	0.31	-0.17
	CE	16.37 ± 0.30	14.14 ± 0.51	12.70-20.70	16.35	1.41	8.62	0.15	-0.34

[#]Seed length (SL), seed width (SW), seed thickness (ST) and 100-seed weight (HSW).
[a]Environment.
[b]Parents of GB13 RIL population.
[c]Female parent of GB13 RIL population.
[d]Male parent of GB13 RIL population.
[e]Recombinant inbred lines.
[f]Range of seed size and 100-seed weight for GB13 RIL population.
[g]Mean of seed size and 100-seed weight for GB13 RIL population.
[h]Standard deviation.
[i]Coefficient of variation.
[j]Skewness.
[k]Kurtosis.

3.3 Analysis of variance and estimates of broad-sense heritability

The results of ANOVA for SL, SW, ST and HSW of the GB13 RIL population in five natural environments (Table 2) show that the genotype, environment and interaction between genotype and environment had significant effects on seed size and weight traits of the GB13 RIL population (p<0.0001). The h^2 for seed size and weight traits of the GB13 RIL population were relatively high ranging from 0.74 to

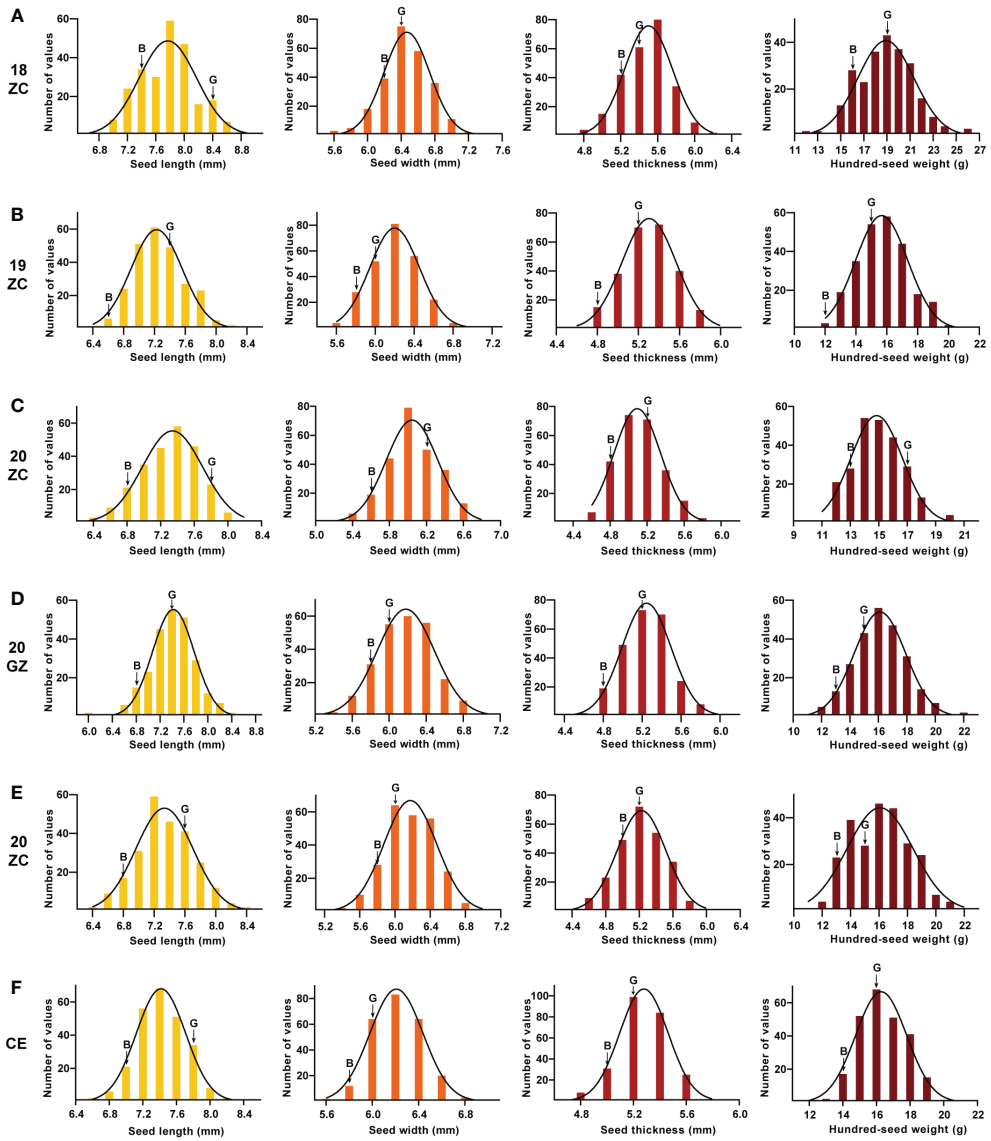


FIGURE 3
Frequency distribution and its fitted curve ofseed size and weight traits for GB13 RIL population across the Teaching and Research Base of Zengcheng (ZC) in **(A)** 2018, **(B)** 2019 and **(C)** 2020, the Experimental Teaching Base of Guangzhou (GZ) in **(D)** 2019 and **(E)** 2020 and **(F)** their combined environment (CE). The letter "G" indicates the female parent Guizao1, and the letter "B" indicates the male parent B13. The horizontal coordinates of the different colored figures from left to right were seed length, seed width, seed thickness and 100-seed weight.

0.83 (Table 2). The traits are suitable for further QTL localization analysis.

In terms of the proportion of total variation accounted for variation of different sources (Table 2), the largest source of variation in SL and SW was genotype, while the largest source of variation in ST and HSW was genotype × environment interactions. The most influenced by the environment was HSW, accounted for 30.71% of the total variation. It was followed by SL, SW and ST, which accounting for 19.51%, 18.19% and 17.53% of the total variation, respectively. The above results indicated that the seed weight trait was more influenced by the environment compared to the seed size trait.

3.4 Identification of stable QTL for seed size and weight traits

A total of 85 QTL were detected for seed size and weight in six environments with PVE in the range of 3.14% - 21.93%, of which 19 QTLs were identified for SL, 23 for SW, 18 for ST, and 25 for HSW (Supplementary Tables 1–4). By collation and

10.3389/fpls.2022.1094112

TABLE 2 Analysis of variance and broad-sense heritability for GB13 RIL population at five natural environments.

Trait[#]	Sources[a]	*Df* [b]	*SS* [c]	*MS* [d]	*P* value[e]	VC[f]	PV (%)[g]	h^2 (%)[h]
SL	Repeat	2	0.74	0.37	<0.0001			0.83
	Genotype	247	294.11	1.19	<0.0001	0.07	46.49	
	Environment	4	123.42	30.86	<0.0001	0.04	19.51	
	Interaction	988	200.95	0.20	<0.0001	0.07	31.77	
	Error	2478	13.38	0.01				
	Total variation	3719	632.59					
SW	Repeat	2	1.09	0.54	<0.0001			0.82
	Genotype	247	164.87	0.67	<0.0001	0.04	45.42	
	Environment	4	66.05	16.51	<0.0001	0.02	18.19	
	Interaction	988	119.22	0.12	<0.0001	0.04	32.84	
	Error	2478	11.80	0.01				
	Total variation	3719	363.04					
ST	Repeat	2	1.09	0.54	<0.0001			0.74
	Genotype	247	118.51	0.48	<0.0001	0.02	38.23	
	Environment	4	54.34	13.58	<0.0001	0.02	17.53	
	Interaction	988	124.30	0.13	<0.0001	0.04	40.1	
	Error	2478	11.72	0.01				
	Total variation	3719	309.96					
HSW	Repeat	2	45.00	22.50	<0.0001			0.75
	Genotype	247	7210.17	29.19	<0.0001	1.46	33.32	
	Environment	4	6645.42	1661.36	<0.0001	2.22	30.71	
	Interaction	988	7256.28	7.34	<0.0001	2.38	33.54	
	Error	2478	479.53	0.19				
	Total variation	3719	21636.40					

[#] Seed length (SL), seed width (SW), seed thickness (ST) and 100-seed weight (HSW).
[a]Sources of variation.
[b]Degree of freedom.
[c]Sum of deviation squares.
[d]mean square.
[e]The P value of the F-test (joint hypotheses test).
[f]Variance components for different sources of variation.
[g]Proportion of variation.
[h]Broad-sense heritability.

analysis, a total of 18 stable QTLs were identified (Table 3), including 3 for SL, 6 for SW, 4 for ST, and 5 for HSW (Figure 4). One QTL, *qSL-3-1*, associated with SL was detected in all environments, and it explained more than 10% of PV in three environments in the range of 11.93% - 15.91%. Three QTLs (*qHSW-11-3*, *qHSW-20-3* and *qSW-20-3*) were detected in four environments. Among them, *qSW-20-3* explained 9.22%, 10.79%, 20.89% and 21.93% of PV in four environments, respectively. Another 14 QTLs distributed on chromosomes 03, 04, 05, 07, 11, 12, 13, 18 and 20 were detected in three environments. In summary, integrating the PVE for QTL, the number of detections in various environments, and the relevant references, *qSL-3-1* and *qSW-20-3* can be considered as novel stable QTL with the main effects. Meanwhile, these stable QTLs above were important references for the excavation of genes controlling seed size and weight traits.

TABLE 3 Stable QTLs associated withseed size and weight traits for GB13 RIL population across different environments.

QTL[#]	Chr[a]	Interval	Position (cM)	PVE (%)[h]						LOD[i]						ADD[j]					
				E1[b]	E2[c]	E3[d]	E4[e]	E5[f]	E6[g]	E1	E2	E3	E4	E5	E6	E1	E2	E3	E4	E5	E6
qSL-3-1	*3*	bin32-bin43	43.1	13.57	6.03	11.93	4.38	6.61	15.91	9.37	4.03	7.49	3.15	4.61	11.45	0.15	0.09	0.11	0.08	0.09	0.11
qSL-3-2	*3*	bin51-bin55	49.4			12.15		9.07	14.56			7.63		6.42	10.38			0.11		0.11	0.11
qHSW-3-2	*3*	bin53-bin55	50.1			5.12		4.41	6.81			3.68		3.20	5.43			0.38		0.37	0.37
qHSW-4-2	*4*	bin75-bin78	57.5	5.34		4.68			5.28	3.75		3.25			4.18	0.59		0.37			0.33
qST-4-3	*4*	bin77-bin83	65.2	4.86	8.61				6.10	3.18	5.44				4.70	0.06	0.08				0.04
qST-5-2	*5*	bin155-bin158	135.1	4.67			4.52		3.69	3.06			3.13		2.89	-0.06			-0.06		-0.04
qST-7-1	*7*	bin159-bin161	139.9				3.85	5.75	6.99				2.69	4.06	5.37				0.05	0.06	0.05
qSW-11-1	*11*	bin95-bin100	101	3.89			7.48		6.34	2.90			5.23		5.14	-0.06			-0.08		-0.05
qHSW-11-3	*11*	bin131-bin134	133.1		4.90		8.26	3.33	4.41		2.99		5.76	2.56	3.41		0.42		0.59	0.32	0.31
qSW-12-2	*12*	bin103-bin109	111.1			5.79		3.69	5.39			4.37		2.83	4.36			0.06		0.05	0.05
qHSW-13-1	*13*	bin115-bin116	90.0		4.34			5.19	7.21		2.91			3.79	5.69		0.39			0.40	0.38
qSL-13-1	*13*	bin115-bin117	90	3.47				4.12	4.47	2.55				3.03	3.54	0.08				0.07	0.06
qST-13-3	*13*	bin143-bin149	104.6		5.09	5.52			4.56		3.27	3.74			3.50		0.06	0.06			0.04
qSW-18-1	*18*	bin237-bin239	152.4	5.62		4.91			3.91	4.03		3.72			3.24	0.07		0.05			0.04
qSW-20-2	*20*	bin46-bin49	49.1	7.46	11.80	8.64				5.30	8.29	6.24				-0.08	-0.10	-0.07			
qSW-20-3	*20*	bin64-bin68	55.3	9.22		10.79		21.93	20.89	6.63		7.89		15.04	15.24	-0.09	-0.10	-0.08		-0.14	-0.10
qHSW-20-3	*20*	bin68-bin73	55.9		6.12	3.31		4.68	7.21		4.06	2.52		3.42	5.70		-0.47	-0.31		-0.38	-0.38
qSW-20-4	*20*	bin84-bin86	64.2			3.80		7.30	5.81			2.65		4.56	3.86			-0.05		-0.09	-0.05

[#]The name of each QTL. [a]Chromosome. [b]at Zengcheng in 2018. [c]at Guangzhou in 2019. [d]at Zengcheng in 2019. [e]at Guangzhou in 2020. [f]at Zengcheng in 2020. [g]Combined environment. [h]Phenotypic variation explained. [i]Log of odd value. [j]Additive effect.

10.3389/fpls.2022.1094112

10.3389/fpls.2022.1094112

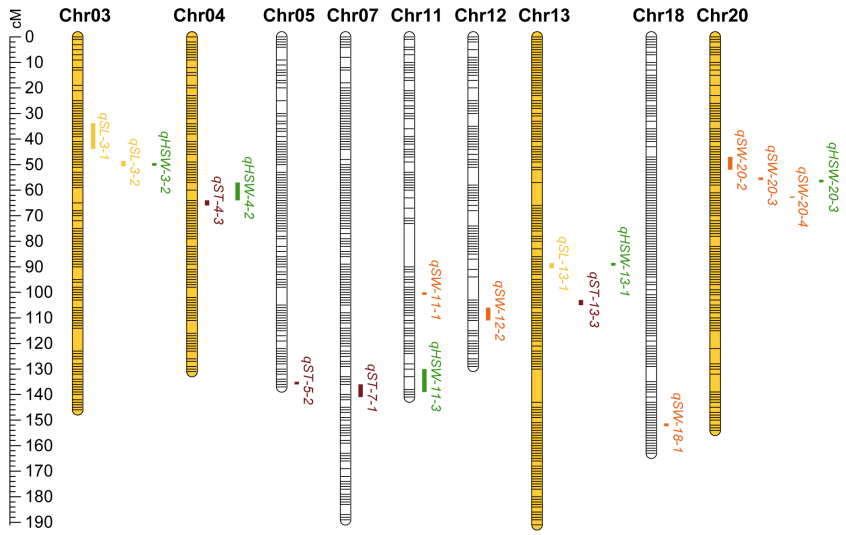


FIGURE 4
The linkage map of stable QTLs for soybean seed size and weight traits. The ruler on the left side indicates the interval distance between markers using cM (centimorgan) as the unit. Bars with yellow emphasis indicate the chromosome on which the QTL clusters are located. Between the bars in the graph, the yellow indicates QTL for SL; the orange indicates QTL for SW; the reddish brown indicates QTL for ST; the green indicates QTL for HSW.

3.5 Identification of QTL clusters

According to the classification criteria of QTL clusters, we obtained five QTL clusters (Table 4) from the QTL mapping results (Supplementary Tables 1–4). These five QTL clusters were distributed on chromosomes 03, 04, 13 and 20. All of them were associated with HSW, which verified that the seed size was closely related to HSW from another aspect. The four QTL clusters related to SL were *qLH3-1*, *qLH3-2*, *qLWTH4* and *qLH13*; the two QTL clusters related to SW were *qLWTP4* and *qWTH20*; the two QTL clusters related to ST were *qLWTP4* and *qWTP20*. In terms of the number of traits controlled, one QTL cluster for four traits was *qLWTP4*, one QTL cluster for three traits was *qWTH20*, and three QTL clusters for two traits were *qLH3-1*, *qLH3-2*, and *qLH13*. Among them, *qLP3-1* and *qLP3-2* containing major and stable QTL(s) were located on the forearm

TABLE 4 QTL clusters associated with at least two traits.

Clusters	Chr[a]	Interval	Position(bp)	QTLs[b]	ADD[c]	PVE(%)[d]	Env.[e]
qLH3-1	3	bin32-bin43	4387438-6456915	*qSL-3-1*	0.15	13.57	18ZC
					0.09	6.03	19GZ
					0.11	11.93	19ZC
					0.08	4.38	20GZ
					0.09	6.61	20ZC
					0.11	15.91	CE
				qHSW-3-1	0.33	3.77	19ZC
					0.35	6.07	CE
qLH3-2	3	bin51-bin55	8784766-17199876	*qSL-3-2*	0.11	12.15	19ZC
					0.11	9.07	20ZC
					0.11	14.56	CE
				qHSW-3-2	0.38	5.12	19ZC

(Continued)

10.3389/fpls.2022.1094112

TABLE 4 Continued

Clusters	Chr[a]	Interval	Position(bp)	QTLs[b]	ADD[c]	PVE(%)[d]	Env.[e]
					0.37	4.41	20ZC
					0.37	6.81	CE
qLWTH4	4	bin74-bin83	9040807-12316836	*qSL-4-2*	0.06	4.36	CE
				qSW-4-2	0.06	3.78	18ZC
					0.04	3.70	CE
				qST-4-3	0.06	4.86	18ZC
					0.08	8.61	19GZ
					0.04	6.10	CE
				qHSW-4-2	0.59	5.34	18ZC
					0.37	4.68	19ZC
					0.33	5.28	CE
qLH13	13	bin115-bin117	27473391-27671175	*qSL-13-1*	0.08	3.47	18ZC
					0.07	4.12	20ZC
					0.06	4.47	CE
				qHSW-13-1	0.39	4.34	19GZ
					0.40	5.19	20ZC
					0.38	7.21	CE
qWTH20	20	bin64-bin73	27890104-33222868	*qSW-20-3*	-0.09	9.22	18ZC
					-0.08	10.79	19ZC
					-0.14	21.93	20ZC
					-0.10	20.89	CE
				qST-20-2	-0.05	7.55	CE
				qHSW-20-3	-0.47	6.12	19GZ
					-0.31	3.31	19ZC
					-0.38	4.68	20ZC
					-0.38	7.21	CE

[a]Chromosome. [b]Indivadual QTLs of cluster. [c]Additive effect. [d]Phenotypic variation explained. [e]Environment.

of chromosome 03 at the physical positions between 4387438 and 6456915 bp and between 8784766 and 17199876 bp, respectively. Similarly, *qWTH20* also contained major and stable QTLs and was located at the physical position of chromosome 20 between 12056958 and 33222868 bp.

In summary, *qLP3-1*, *qLP3-2* and *qWTH20* may be QTL clusters containing major effect genes that regulate soybean seed size and weight traits; *qLWTP4* and *qLP13* may be QTL clusters containing micro-effector genes that harmonize the control of soybean seed size and weight traits. Therefore, these five QTL clusters can provide a reference for mining the target genes controlling seed size and weight traits, and were used as the target intervals for gene GO enrichment analysis.

3.6 Gene GO enrichment analysis

By Gene GO enrichment analysis, we found that most of the genes within these five QTL cluster intervals were involved in cellular processes and metabolic processes (Figure 5). Most of the genes in the *qLP3-1* were involved in intracellular metabolic processes, bioregulation and reaction to stimuli, and were mostly involved in binding and catalytic activities in terms of molecular

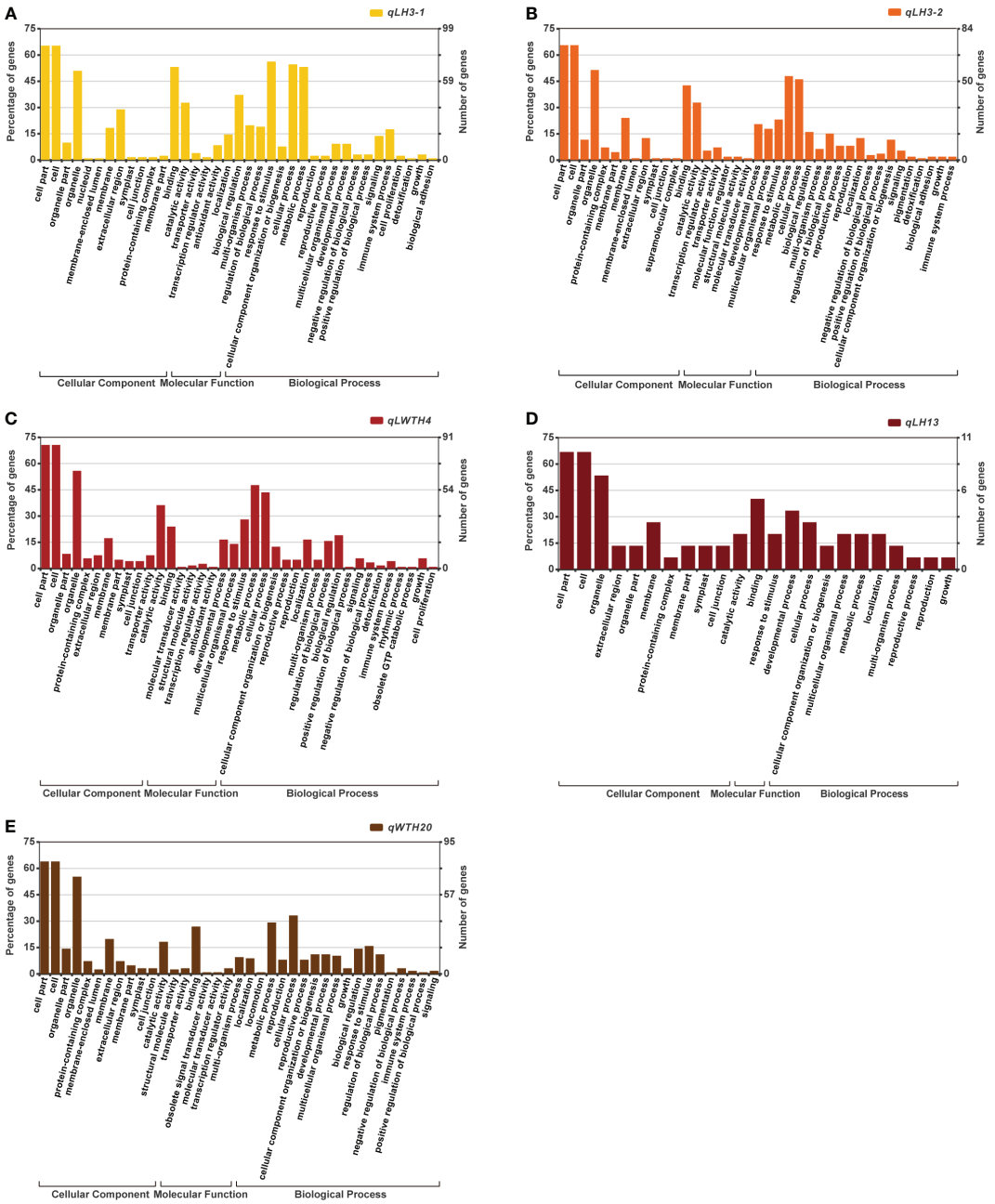


FIGURE 5
GO term enrichment analysis of the genes located within the five QTL clusters: **(A)** *qLH3-1*; **(B)** *qLH3-2*; **(C)** *qLWTH4*; **(D)** *qLH13*; **(E)** *qWTH20*.

functions. While most of the genes in the *qLP3-2* were similar in function to those in the *qLP3-1*. In comparison, most of the genes in the *qLWTP4* were also involved in intracellular metabolic processes, but a small number of genes were involved in regulating growth and cell proliferation. The *qLP13* contained only 11 annotated genes, most of which were involved in growth and cellular processes. Similarly, most of the genes in the *qWTP20-2* were involved in intracellular metabolic processes.

3.7 Candidate genes in QTL clusters

Through the above gene GO enrichment analysis, followed by gene expression screening, gene function annotation, and sequence variation analysis, we obtained a total of 15 candidate genes (Table 5) that might be participating in the regulation of seed size and weight traits. Among them, six genes were located on chromosome 03, three on chromosome 04, two on chromosome 13, and four on chromosome 20. Among the 15

TABLE 5 Candidate genes identified within the five QTL clusters that were highly expressed or specific expression in soybean seed.

Gene	Position (bp)	Gene functional annotation
Glyma.03g036600	4432910-4435002	Negative regulation of transcription; organ development; lipid transport calmodulin binding
Glyma.03g040400	5050463-5052307	Protease inhibitors (LTP family); Seed storage
Glyma.03g044200	5586338-5589780	Transporter protein activity; integral to membrane
Glyma.03g045400	5760642-5763461	Metabolic process; methyltransferase activity
Glyma.03g065700	11896538-11909845	Regulation of transcription; sequence-specific DNA binding transcription factor activity; DNA-dependent
Glyma.03g065900	12064850-12067709	Metabolic process; hydrolase activity; catalytic activity
Glyma.04g100100	9177207-9180239	Negative regulation of gene expression; response to abscisic acid stimulus; ubiquitin-protein ligase activity
Glyma.04g100400	9219564-9221412	Sterol biosynthetic process; response to oleuropein lactone stimulus
Glyma.04g107100	11239281-11250948	Regulation of meristem growth; Ubiquitin ligases are involved in synaptic protein degradation
Glyma.13g159500	27476780-27479031	Ubiquitin system components suggestive of protein; cell growth; cell morphogenesis
Glyma.13g160400	27606020-27608359	Lipid transport; lipid binding; Hydrophobic protein
Glyma.20g081600	30778586-30779164	Regulation of gene expression; response to auxin stimulus; protein binding
Glyma.20g084000	31536334-31538924	RNA methylation; RNA processing
Glyma.20g084500	31666190-31674959	Ubiquitin-protein ligase activity; protein ubiquitination; CUL4-RING ubiquitin ligase complex
Glyma.20g087000	32527435-32530216	Negative regulation of ethylene mediated signaling pathway; regulation of transcription; ethylene binding

genes screened, *Glyma.03g040400* was the gene with the same function as the gene (*LOC_Os07g19000*) associated with seed size in rice; *Glyma.03g045400* and *Glyma.20g084000* were related to DNA methylation; *Glyma.04g100400*, *Glyma.20g081600* and *Glyma.20g087000* were hormone-related and might be involved in the regulation of seed size and weight traits by hormone signaling pathway; *Glyma.04g100100*, *Glyma.04g107100*, *Glyma.13g159500* and *Glyma.20g084500* might negatively regulate seed size and weight through the ubiquitin-proteasome pathway. *Glyma.03g036600* and *Glyma.03g065700* were associated with transcription factor regulation and might be involved in the regulation of seed size and weight traits through the transcription factor regulation pathway; *Glyma.03g044200*, *Glyma.03g065900* and *Glyma.13g160400* might be related to the metabolic synthesis of protein and oil, and indirectly regulated the seed size and weight traits.

The expression heat map (Figure 6A) of these 15 candidate genes showed that two genes each were highly expressed in young leaf and nodule of soybean; three genes were specifically expressed at the pod shell development stage in soybean; eight genes including *Glyma.03g036600*, *Glyma.03g040400*, *Glyma.03g045400*, *Glyma.03g065700*, *Glyma.03g065900*, *Glyma.04g100400*, *Glyma.04g107100* and *Glyma.20g087000* were specifically and progressively expressed at the seed maturity stage in soybean. Meanwhile, the structural maps for most of the candidate genes (Figure 6B) were matched with their sequence in the parental lines, but ten candidate genes varied in the sequence of the parental lines, with the variant region on introns, 3'UTR (untranslated region) and 5'UTR (Table 6). They are *Glyma.03g036600*, *Glyma.03g040400*, *Glyma.03g044200*, *Glyma.03g045400*, *Glyma.03g065700*, *Glyma.04g100100*, *Glyma.04g107100*, *Glyma.13g160400*, *Glyma.20g084000* and *Glyma.20g084500*, and deserve to be investigated in depth.

4 Discussion

Based on the results of the correlation analysis between seed size and seed weight in this study, the R2 between seed width and seed weight was the largest in most environments, so we suggest that varieties with wide seeds can be selected for varietal improvement for soybean seed weight trait. However, from the results of the QTLs localized for soybean seed size and weight traits, most of the QTLs for HSW overlapped or were close to the QTL interval for SL. It suggests that most of the genes with multi-effects may be genes that coordinate the control on both SL and HSW. Nevertheless, the specific regulatory mechanism is still unclear (He et al., 2021) and needs to be further investigated.

In this study, a total of 60 QTLs related to seed size trait and 25 QTLs related to seed weight trait were identified in the GB13 RIL population in five natural environments and their combined environment. Nineteen QTLs for SL included 4 stable QTLs and 15 sensitive QTLs (Supplementary Table 1); 23 QTLs for seed width included 6 stable QTLs and 17 sensitive QTLs (Supplementary Table 2); 18 QTLs for ST included 5 stable QTLs and 13 sensitive QTLs (Supplementary Table 3). The 25 QTLs associated with HSW included 5 stable QTLs and 20 sensitive QTLs (Supplementary Table 4). By comparing the

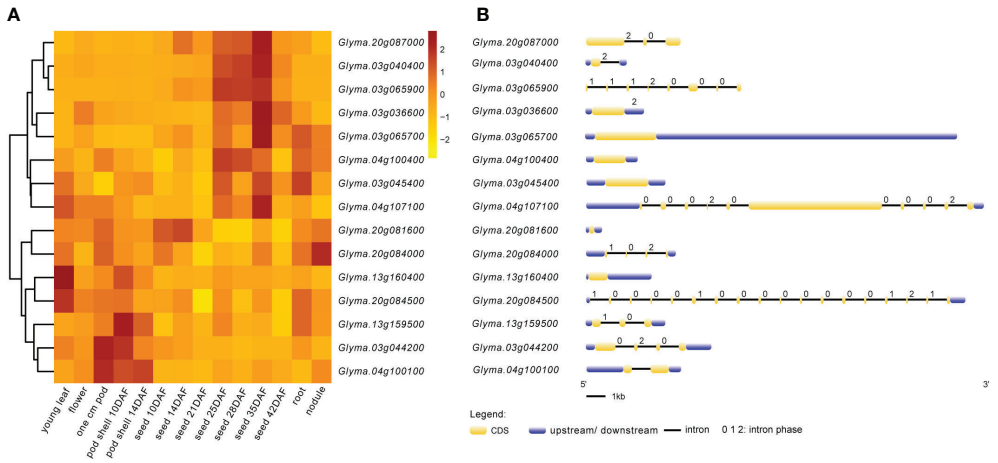


FIGURE 6
Expression patterns and structures of candidate genes. **(A)** Visual heat map of candidate genes expressed in different tissues and periods of soybean. The expression data of 15 candidate genes in this study were downloaded from SoyBase (https://www.soybase.org/soyseq/). Normalization of expression data by different genes. **(B)** Gene structure diagram of candidate genes. The yellow indicates CDS (coding DNA sequence); the blue indicates upstream or downstream of CDS containing 5'UTR and 3'UTR. Arabic numbers indicate the number of intron phases.

results with those of previous studies, ten sensitive QTLs for SL (*qSL-2-1*, *qSL-3-3*, *qSL-4-1*, *qSL-4-2*, *qSL-4-3*, *qSL-6-2*, *qSL-11-1*, *qSL-11-2*, *qSL-18-2* and *qSL-20-1*) overlapped with the intervals of previous localization results or were within their subsets (Salas et al., 2006; Xu et al., 2011; Niu et al., 2013; Yang et al., 2017; Cui et al., 2020; Hina et al., 2020; Kumawat and Xu, 2021). Eight sensitive QTLs for SW (*qSW-2-1*, *qSW-2-2*, *qSW-4-1*, *qSW-9-1*, *qSW-10-1*, *qSW-15-1*, *qSW-20-1*, and *qSW-20-5*) overlap with or were within a subset of the previous localization result intervals (Xu et al., 2011; Niu et al., 2013; Yang et al., 2017; Hina et al., 2020). Five sensitive QTLs for ST (*qST-4-4*, *qST-5-2*, *qST-6-1*, *qST-13-1* and *qST-20-3*) overlapped with or included the intervals of previous localization results (Xu et al., 2011; Niu et al., 2013; Hina et al., 2020). Five sensitive QTLs for HSW (*qHSW-4-3*, *qHSW-13-2*, *qHSW-17-2*, *qHSW-20-1*, and *qHSW-20-5*) were found to have regions of overlap with the localization results of previous studies (Mansur et al., 1993; Teng et al., 2009; Han et al., 2012; Kato et al., 2014). Overall, the results of the present study correlated well with those of previous studies.

Besides, comparing the localization results of seed size traits in this study with the results of other related trait localization studies, eight QTLs (*qSL-4-1*, *qSL-17-1*, *qSL-20-1*, *qSW-12-2*, *qSW-15-2*, *qSW-17-1*, *qST-13-3*, *qST-20-1*) were found to overlap or interval overlap with QTLs related to yield traits from previous studies (Mansur et al., 1993; Kabelka et al., 2004; Du et al., 2009; Palomeque et al., 2009; Liu et al., 2011; Wang X. et al., 2014). Fifteen QTLs (*qSL-6-2*, *qSL-13-1*, *qSL-17-1*, *qSL-18-1*, *qSL-20-1*, *qSW-11-1*, *qSW-12-1*, *qSW-12-2*, *qSW-20-2*, *qSW-20-3*, *qST-5-2*, *qST-13-1*, *qST-13-3*, *qST20-1*, *qST-20-2*) overlapped QTLs associated with traits related to protein, oil and fatty acids (Sebolt et al., 2000; Tajuddin et al., 2003; Bachlava et al., 2009; Qi et al., 2011; Eskandari et al., 2013; Mao et al., 2013; Wang J. et al., 2014; Fan et al., 2015; Warrington et al., 2015; Han et al., 2015; Leite et al., 2016; Fliege et al., 2022). Five QTLs (*qSW-15-1*, *qSW-15-3*, *qSW-18-2*, *qST-4-1* and *qST-5-2*) had overlapping intervals with QTLs related to seed-set of pod number and soybean maturity (Yao et al., 2015; Ning et al., 2018). Two QTLs (*qSL-5-1* and *qST-5-2*) had interval overlapping with isoflavonoid-related QTLs (Yang et al., 2011; Cai et al., 2018). The *qSW-4-1* overlapped with the QTLs associated with the long juvenility (Yue et al., 2017). Researchers sometimes studied the seed weight of soybean for QTLs mapping in collaboration with protein and oil content (Fasoula et al., 2004; Pathan et al., 2013; Yesudas et al., 2013; Wang X. et al., 2014; Whiting et al., 2020; He et al., 2021). In this study, *qHSW-13-1*, *qHSW-15-1*, *qHSW-17-2*, *qHSW-20-2* and the stable QTL *qHSW-20-3* had overlap intervals with QTLs related to protein or oil content studied previously (Panthee et al., 2005; Mao et al., 2013; Wang X. et al., 2014; Han et al., 2015; Fliege et al., 2022). Besides, *qHSW-15-1* and *qHSW-17-1* also had overlap intervals with yield-related QTL studied previously (Palomeque et al., 2009; Liu et al., 2011). In summary, many QTLs with stable main effects related to seed size and weight traits in this study were consistent with previous studies results and correlated with QTLs related to other traits including protein, oil content, pod number, yield, maturity, and long juvenile stage. Thus, we suggest that genes regulating soybean seed size and weight traits may be correlated with genes regulating other traits, especially protein and oil content, so genes regulating the synthesis or metabolism of protein and oil content were also used as the basis for candidate gene speculation in this study.

In addition, compared with soybean, rice (*Oryza sativa L.*) has been studied much more intensively than soybean in terms of its seed size and weight. Currently, genes associated with seed size and weight, such as *GW2* (Song et al., 2007), *OsGRF4* (Li et al.,

10.3389/fpls.2022.1094112

TABLE 6 Sequence variations between the parental lines of candidate genes.

Gene	Chr[a]	Loci(bp)[b]	Guizao1[c]	B13[d]	Region[e]
Glyma.03g036600	3	4433324	T	TA	UTR3
Glyma.03g040400	3	5051148	A	ACG	intronic
	3	5051173	A	ATCACC	intronic
	3	5051386	T	TC	intronic
	3	5052301	G	GAT	UTR3
Glyma.03g044200	3	5588615	TCA	T	intronic
	3	5588681	AAAC	A	intronic
	3	5588699	AAC	A	intronic
Glyma.03g045400	3	5760781	T	TTC	UTR3
	3	5760989	TG	T	UTR3
	3	5761009	A	AAAAACAAGAG	UTR3
	3	5761114	AACTCTC	A	UTR3
	3	5763119	T	TATG	UTR5
Glyma.03g065700	3	11907032	G	GA	UTR3
Glyma.03g065900	3	NA	NA	NA	NA
Glyma.04g100100	4	9179539	G	GAAAACAAAAC	UTR5
Glyma.04g100400	4	NA	NA	NA	NA
Glyma.04g107100	4	11246102	GA	G	intronic
Glyma.13g159500	13	NA	NA	NA	NA
Glyma.13g160400	13	27607021	TA	T	UTR3
Glyma.20g081600	20	NA	NA	NA	NA
Glyma.20g084000	20	31536882	A	AAAAACCCTACATCTTCAT	intronic
	20	31537182	ATT	A	intronic
	20	31537333	TTA	T	intronic
	20	31537718	A	AGAATCAATGTCATT	intronic
	20	31538735	TTA	T	UTR3
Glyma.20g084500	20	31666328	GTCA	G	UTR3
	20	31669438	A	AAC	intronic
	20	31673766	TCG	T	intronic
	20	31674346	T	TAA	intronic
	20	31674567	T	TA	intronic
Glyma.20g087000	20	NA	NA	NA	NA

[a]Chromosome.
[b]Positions with variation in genes between parental lines.
[c]Female parent of GB13 RIL population.
[d]Male parent of GB13 RIL population.
[e]Region of variation in genes.

2016), *GS3* (Fan et al., 2006), *qSW5* (Shomura et al., 2008), *GW5* (Weng et al., 2008), *qGL3* (Zhang et al., 2012), *LGY3* (Liu Q. et al., 2018), *GS5* (Li and Li, 2016), *GS6* (Sun et al., 2013), *GLW7* (Wang et al., 2015), etc., have been cloned. Furthermore, the regulatory pathways of the genes mainly include the ubiquitin-proteasome pathway, G protein signaling pathway, Mitogen-activated protein kinases (MAPK) signaling pathway, hormone signaling pathway and transcription factor regulatory pathway, etc. (Li and Li, 2016). Among them, in the ubiquitin-proteasome pathway, E3 ubiquitin ligase plays a major role and negatively regulates seed width and seed weight (Song et al., 2007). In the hormone signaling pathway, Brassinolide (BR) and Auxin (IAA) can promote cell growth and expansion (Li et al., 2019), it mainly regulates glume development to positively regulate seed size in rice (Li et al., 2011). Therefore, in this study, candidate genes were selected and predicted by screening genes within five clusters based on their specific expression in soybean, combined with homologous or identical functions of genes related to grain size and weight traits in rice.

The major and stable QTL, *qSW-20-3*, which was detected in four environments with PVE varied from 9.22 to 21.93%, was contained in the *qWTP20-2* cluster. By the analysis of sequence variation, we found that among the candidate genes within this QTL interval, only *Glyma.20g084000* and *Glyma.20g084500* were different in the sequence of the parental lines. They were not specifically expressed at the seed maturation stage in soybean, moreover, they have a negative additive effect, but they were associated with RNA methylation and protein ubiquitination, respectively. Therefore, we speculate that the genes within this cluster may indirectly regulate seed size and weight by regulating other traits. While another major and stable QTL, *qSL-3-1*, which was detected in six environments with PVE varied from 4.38 to 15.91%, was contained in the *qLP3-1* cluster. Comparing the sequence variation between parental lines for candidate genes within this cluster, we found that five genes underwent natural variation between parental lines, and all of them were specifically and highly expressed at the seed maturation stage in soybean. Moreover, they have a positive additive effect. Therefore, we speculate that the five candidate genes within the *qLP3-1* cluster regulate seed size and weight positively through certain pathways based on their gene annotation. However, the exact signaling pathway remains to be further investigated.

It is important to study the molecular regulatory network of yield-related traits in soybean. And the loci and candidate genes in this study provide an important theoretical basis and genetic resources for solving the bottleneck problem of soybean yield, which deserves to be further investigated at the molecular level.

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

YC and SL designed the project. SL performed the experiments and drafted the manuscript. RL, ZG, BC, and FL helped the experiment to obtain part of the data. JJ and RW revised the manuscript. SL and QX analyzed the data. HN and ZC validated the manuscripts.

Funding

This work was supported by the Key-Areas Research and Development Program of Guangdong Province (2020B020220008), the China Agricultural Research System (CARS-04-PS09), the Guangdong Agricultural Research System (2022KJ136), and the Research Project of the State Key Laboratory of Agricultural and Biological Resources Protection and Utilization in Subtropics.

Acknowledgments

We thank Prof. Liangfa Ge (South China Agricultural University, Guangzhou, China) for critical reading and comments on the manuscript and Long Xiao (Perkins&Will, Vancouver, BC) for constructive suggestions on the language of the manuscript.

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.1094112/full#supplementary-material

References

Bachlava, E., Dewey, R. E., Burton, J. W., and Cardinal, A. J. (2009). Mapping and comparison of quantitative trait loci for oleic acid seed content in two segregating soybean populations. *Crop Sci.* 49 (2), 433–442. doi: 10.2135/cropsci2008.06.0324

Cai, Z., Cheng, Y., Ma, Z., Liu, X., Ma, Q., Xia, Q., et al. (2018). Fine-mapping of QTLs for individual and total isoflavone content in soybean (*Glycine max* l.) using a high-density genetic map. *Theor. Appl. Genet.* 131 (3), 555–568. doi: 10.1007/s00122-017-3018-x

Chapman, A., Pantalone, V. R., Ustun, A., Allen, F. L., Landau-Ellis, D., Trigiano, R. N., et al. (2003). Quantitative trait loci for agronomic and seed quality traits in an F_2 and $F_{4:6}$ soybean population. *Euphytica.* 129 (3), 387–393. doi: 10.1023/A:1022282726117

Csanadi, 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 (6-7), 912–919. doi: 10.1007/s001220100621

Cui, B., Chen, L., Yang, Y., and Liao, H. (2020). Genetic analysis and map-based delimitation of a major locusqSS3 for seed size in soybean. *Plant Breeding.* 139 (6), 1145–1157. doi: 10.1111/pbr.12853

Du, W., Wang, M., Fu, S., and Yu, D. (2009). Mapping QTLs for seed yield and drought susceptibility index in soybean (*Glycine max* l.) across different environments. *J. Genet. Genomics* 36 (12), 721–731. doi: 10.1016/S1673-8527(08)60165-4

Elattar, M. A., Karikari, B., Li, S., Song, S., Cao, Y., Aslam, M., et al. (2021). Identification and validation of major QTLs, epistatic interactions, and candidate genes for soybean seed shape and weight using two related RIL populations. *Front. Genet.* 12. doi: 10.3389/fgene.2021.666440

Eskandari, M., Cober, E. R., and Rajcan, I. (2013). Genetic control of soybean seed oil: I. QTL and genes associated with seed oil concentration in RIL populations derived from crossing moderately high-oil parents. *Theor. Appl. Genet.* 126 (2), 483–495. doi: 10.1007/s00122-012-1995-3

Fan, S., Li, B., Yu, F., Han, F., Yan, S., Wang, L., et al. (2015). Analysis of additive and epistatic quantitative trait loci underlying fatty acid concentrations in soybean seeds across multiple environments. *Euphytica.* 206 (3), 689–700. doi: 10.1007/s10681-015-1491-3

Fan, C., Xing, Y., Mao, H., Lu, T., Han, B., Xu, C., et al. (2006). GS3, a major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein. *Theor. Appl. Genet.* 112 (6), 1164–1171. doi: 10.1007/s00122-006-0218-1

Fasoula, V. A., Harris, D. K., and Boerma, H. R. (2004). Validation and designation of quantitative trait loci for seed protein, seed oil, and seed weight from two soybean populations. *Crop Sci.* 44 (4), 1218–1225. doi: 10.2135/cropsci2004.1218

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 quantitative trait loci on soybean chromosome 20. *Plant J.* 110 (1), 114–128. doi: 10.1111/tpj.15658

Gao, H. H., Gu, Y., Jiang, H. W., Li, Y. Y., Xin, D. W., Liu, C. Y., et al. (2019). Mapping and analysis of QTLs related to seed length and seed width in *Glycine max*. *Plant Breeding.* 138 (6), 733–740. doi: 10.1111/pbr.12745

Han, Y., Teng, W., Wang, Y., Zhao, X., Wu, L., Li, D., et al. (2015). Unconditional and conditional QTL underlying the genetic interrelationships between soybean seed isoflavone, and protein or oil contents. *Plant Breeding.* 134 (3), 300–309. doi: 10.1111/pbr.12259

Han, Y., Xie, D., Teng, W., Sun, J., and Li, W. (2012). QTL underlying developmental behaviour of 100-seed weight of soybean. *Plant Breeding.* 131 (5), 600–606. doi: 10.1111/j.1439-0523.2012.01987.x

He, Q., Xiang, S., Yang, H., Wang, W., Shu, Y., Li, Z., et al. (2021). A genome-wide association study of seed size, protein content, and oil content using a natural population of sichuan and chongqing soybean. *Euphytica.* 217 (11), 198. doi: 10.1007/s10681-021-02931-8

Hina, A., Cao, Y., Song, S., Li, S., Sharmin, R. A., Elattar, M. A., et al. (2020). High-resolution mapping in two RIL populations refines major QTL "hotspot" regions for seed size and shape in soybean (*Glycine max* l.). *Int. J. Mol. Sci.* 21 (3), 1040. doi: 10.3390/ijms21031040

Hu, Z., Zhang, H., Kan, G., Ma, D., Zhang, D., Shi, G., et al. (2013). Determination of the genetic architecture of seed size and shape *via* linkage and association analysis in soybean (*Glycine max* l. merr.). *Genetica.* 141 (4-6), 247–254. doi: 10.1007/s10709-013-9723-8

Hyten, D. L., Pantalone, V. R., Sams, C. E., Saxton, A. M., Landau-Ellis, D., Stefaniak, T. R., et al. (2004). Seed quality QTL in a prominent soybean population. *Theor. Appl. Genet.* 109 (3), 552–561. doi: 10.1007/s00122-004-1661-5

Jiang, B. Z., Li, M., Cheng, Y. B., Cai, Z. D., Ma, Q. B., Jiang, Z., et al. (2019). Genetic mapping of powdery mildew resistance genes in soybean by high-throughput genome-wide sequencing. *Theor. Appl. Genet.* 132 (6), 1833–1845. doi: 10.1007/s00122-019-03319-y

Jia, J., Wang, H., Cai, Z., Wei, R., Huang, J., Xia, Q., et al. (2022). Identification and validation of stable and novel quantitative trait loci for pod shattering in soybean [*Glycine max* (L.) merr.]. *J. Integr. Agr.* 21 (11), 3169–3184. doi: 10.1016/j.jia.2022.08.082

Kabelka, E. A., Diers, B. W., Fehr, W. R., LeRoy, A. R., Baianu, I. C., You, T., et al. (2004). Putative alleles for increased yield from soybean plant introductions. *Crop Sci.* 44 (3), 784. doi: 10.2135/cropsci2004.0784

Kato, S., Sayama, T., Fujii, K., Yumoto, S., Kono, Y., Hwang, T., et al. (2014). A major and stable QTL associated with seed weight in soybean across multiple environments and genetic backgrounds. *Theor. Appl. Genet.* 127 (6), 1365–1374. doi: 10.1007/s00122-014-2304-0

Kulkarni, K. P., Asekova, S., Lee, D., Bilyeu, K., Song, J. T., and Lee, J. (2017). Mapping QTLs for 100-seed weight in an interspecific soybean cross of williams 82 (*Glycine max*) and PI 366121 (*Glycine soja*). *Crop Pasture Sci.* 68 (2), 148–155. doi: 10.1071/CP16246

Kumawat, G., and Xu, D. (2021). A major and stable quantitative trait locus qSS2 for seed size and shape traits in a soybean RIL population. *Front. Genet.* 12. doi: 10.3389/fgene.2021.646102

Lee, S. H., Park, K. Y., Lee, H. S., Park, E. H., and Boerma, H. R. (2001). Genetic mapping of QTLs conditioning soybean sprout yield and quality. *Theor. Appl. Genet.* 103 (5), 702–709. doi: 10.1007/s001220100595

Leite, D. C., Pinheiro, J. B., Campos, J. B., Di Mauro, A. O., and Unêda-Trevisoli, S. H. (2016). QTL mapping of soybean oil content for marker-assisted selection in plant breeding program. *Genet. Mol. Res.* 15 (1), grm7685. doi: 10.4238/gmr.15017685

Li, Y., Fan, C., Xing, Y., Jiang, Y., Luo, L., Sun, L., et al. (2011). Natural variation in GS5 plays an important role in regulating grain size and yield in rice. *Nat. Genet.* 43 (12), 1266–1269. doi: 10.1038/ng.977

Li, S., Gao, F., Xie, K., Zeng, X., Cao, Y., Zeng, J., et al. (2016). The OsmiR396c-OsGRF4-OsGIF1 regulatory module determines grain size and yield in rice. *Plant Biotechnol. J.* 14 (11), 2134–2146. doi: 10.1111/pbi.12569

Li, N., and Li, Y. (2016). Signaling pathways of seed size control in plants. *Curr. Opin. Plant Biol.* 33, 23–32. doi: 10.1016/j.pbi.2016.05.008

Li, M., Liu, Y., Wang, C., Yang, X., Li, D., Zhang, X., et al. (2020). Identification of traits contributing to high and stable yields in different soybean varieties across three chinese latitudes. *Front. Plant Sci.* 10, 1642. doi: 10.3389/fpls.2019.01642

Liu, Q., Han, R., Wu, K., Zhang, J., Ye, Y., Wang, S., et al. (2018). G-Protein βγ subunits determine grain size through interaction with MADS-domain transcription factors in rice. *Nat. Commun.* 9 (1), 852. doi: 10.1038/s41467-018-03047-9

Liu, W., Kim, M. Y., Van, K., Lee, Y., Li, H., Liu, X., et al. (2011). QTL identification of yield-related traits and their association with flowering and maturity in soybean. *J. Crop Sci. Biotechnol.* 14 (1), 65–70. doi: 10.1007/s12892-010-0115-7

Liu, N., Li, M., Hu, X., Ma, Q., Mu, Y., Tan, Z., et al. (2017). Construction of high-density genetic map and QTL mapping of yield-related and two quality traits in soybean RILs population by RAD-sequencing. *BMC Genomics* 18 (1), 466. doi: 10.1186/s12864-017-3854-8

Liu, D., Yan, Y., Fujita, Y., and Xu, D. (2018). Identification and validation of QTLs for 100-seed weight using chromosome segment substitution lines in soybean. *Breed. Sci.* 68 (4), 442–448. doi: 10.1270/jsbbs.17127

Li, N., Xu, R., and Li, Y. (2019). Molecular networks of seed size control in plants. *Annu. Rev. Plant Biol.* 70 (1), 435–463. doi: 10.1146/annurev-arplant-050718-095851

Li, W., Zheng, D., Van, K., and Lee, S. (2008). QTL mapping for major agronomic traits across two years in soybean (*Glycine max* l. merr.). *J. Crop Sci. Biotechnol.* 11 (3), 171–176.

Mansur, L. M., Lark, K. G., Kross, H., and Oliveira, A. (1993). Interval mapping of quantitative trait loci for reproductive, morphological, and seed traits of soybean (*Glycine max* l.). *Theor. Appl. Genet.* 86 (8), 907–913. doi: 10.1007/BF00211040

Mao, T., Jiang, Z., Han, Y., Teng, W., Zhao, X., and Li, W. (2013). Identification of quantitative trait loci underlying seed protein and oil contents of soybean across multi-genetic backgrounds and environments. *Plant Breeding.* 132 (6), 630–641. doi: 10.1111/pbr.12091

Maughan, P. J., Maroof, M. A., and Buss, G. R. (1996). Molecular-marker analysis of seed-weight: Genomic locations, gene action, and evidence for

orthologous evolution among three legume species. *Theor. Appl. Genet.* 93 (4), 574–579. doi: 10.1007/BF00417950

Mccouch, S., Cho, Y., Yano, M., Paul, E., Blinstrub, M., Morishima, H., et al. (1997). Report on QTL nomenclature. *Rice Genet. Newsl.* 14, 11–13.

Mian, M. A., Bailey, M. A., Tamulonis, J. P., Shipe, E. R., Carter, T. E. J., Parrott, W. A., et al. (1996). Molecular markers associated with seed weight in two soybean populations. *Theor. Appl. Genet.* 93 (7), 1011–1016. doi: 10.1007/BF00230118

Ning, H., Yuan, J., Dong, Q., Li, W., Xue, H., Wang, Y., et al. (2018). Identification of QTLs related to the vertical distribution and seed-set of pod number in soybean [*Glycine max* (L.) merri]. *PLoS One* 13 (4), e195830. doi: 10.1371/journal.pone.0195830

Niu, Y., Xu, Y., Liu, X., Yang, S., Wei, S., Xie, F., et al. (2013). Association mapping for seed size and shape traits in soybean cultivars. *Mol. Breeding.* 31 (4), 785–794. doi: 10.1007/s11032-012-9833-5

Nyquist, W. E., and Baker, R. J. (1991). Estimation of heritability and prediction of selection response in plant populations. *Crit. Rev. Plant Sci.* 10 (3), 235–322. doi: 10.1080/07352689109382313

Palomeque, L., Li-Jun, L., Li, W., Hedges, B., Cober, E. R., and Rajcan, I. (2009). QTL in mega-environments: II. agronomic trait QTL co-localized with seed yield QTL detected in a population derived from a cross of high-yielding adapted × high-yielding exotic soybean lines. *Theor. Appl. Genet.* 119 (3), 429–436. doi: 10.1007/s00122-009-1048-8

Panthee, D. R., Pantalone, V. R., West, D. R., Saxton, A. M., and Sams, C. E. (2005). Quantitative trait loci for seed protein and oil concentration, and seed size in soybean. *Crop Sci.* 45 (5), 2015–2022. doi: 10.2135/cropsci2004.0720

Pathan, S. M., Vuong, T., Clark, K., Lee, J., 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 (3), 765–774. doi: 10.2135/cropsci2012.03.0153

Qi, Z., Wu, Q., Han, X., Sun, Y., Du, X., Liu, C., et al. (2011). Soybean oil content QTL mapping and integrating with meta-analysis method for mining genes. *Euphytica.* 179 (3), 499–514. doi: 10.1007/s10681-011-0386-1

Qi, Z. M., Zhang, X. Y., Qi, H. D., Xin, D. W., Han, X., Jiang, H. W., et al. (2017). Identification and validation of major QTLs and epistatic interactions for seed oil content in soybeans under multiple environments based on a high-density map. *Euphytica.* 213 (8), 162. doi: 10.1007/s10681-017-1952-y

Salas, P., Oyarzo-Llaipen, J. C., Wang, D., Chase, K., and Mansur, L. (2006). Genetic mapping of seed shape in three populations of recombinant inbred lines of soybean (*Glycine max* l. merr.). *Theor. Appl. Genet.* 113 (8), 1459–1466. doi: 10.1007/s00122-006-0392-1

Sax, K. (1923). The association of size differences with seed-coat pattern and pigmentation in PHASEOLUS VULGARIS. *Genetics.* 8 (6), 552–560. doi: 10.1093/genetics/8.6.552

Sebolt, A. M., Shoemaker, R. C., and Diers, B. W. (2000). Analysis of a quantitative trait locus allele from wild soybean that increases seed protein concentration in soybean. *Crop Sci.* 40 (5), 1438–1444. doi: 10.2135/cropsci2000.4051438x

Shomura, A., Izawa, T., Ebana, K., Ebitani, T., Kanegae, H., Konishi, S., et al. (2008). Deletion in a gene associated with grain size increased yields during rice domestication. *Nat. Genet.* 40 (8), 1023–1028. doi: 10.1038/ng.169

Song, X., Huang, W., Shi, M., Zhu, M., and Lin, H. (2007). A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase. *Nat. Genet.* 39 (5), 623–630. doi: 10.1038/ng2014

Sun, L., Li, X., Fu, Y., Zhu, Z., Tan, L., Liu, F., et al. (2013). GS6, a member of the GRAS gene family, negatively regulates grain size in rice. *J. Integr. Plant Biol.* 55 (10), 938–949. doi: 10.1111/jipb.12062

Sun, Y. Q., Tian, R., Shao, Z. Q., Chen, S. L., Zhang, H., Jin, Y., et al. (2021). Mining of quantitative trait loci and candidate genes for seed size and shape across multiple environments in soybean (*Glycine max*). *Plant Breeding.* 140 (6), 1058–1069. doi: 10.1111/pbr.12968

Tajuddin, T. C. F. T., Watanabe, S., Yamanaka, N., and Harada, K. (2003). Analysis of quantitative trait loci for protein and lipid contents in soybean seeds using recombinant inbred lines. *Breed. Sci.* 53 (2), 133–140. doi: 10.1270/jsbbs.53.133

Teng, W., Han, Y., Du, Y., Sun, D., Zhang, Z., Qiu, L., et al. (2009). QTL analyses of seed weight during the development of soybean (*Glycine max* l. merr.). *Heredity.* 102 (4), 372–380. doi: 10.1038/hdy.2008.108

Teng, W. L., Sui, M. N., Li, W., Wu, D. P., Zhao, X., Li, H. Y., et al. (2018). Identification of quantitative trait loci underlying seed shape in soybean across multiple environments. *J. Agric. Science.* 156 (1), 3–12. doi: 10.1017/S002185961700082X

Wang, H., Jia, J., Cai, Z., Duan, M., Jiang, Z., Xia, Q., et al. (2022). Identification of quantitative trait loci (QTLs) and candidate genes of seed iron and zinc content in soybean [*Glycine max* (L.) merr.]. *BMC Genomics* 23 (1), 146. doi: 10.1186/s12864-022-08313-1

Wang, X., Jiang, G., 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 (5), 935–949. doi: 10.1007/s00438-014-0865-x

Wang, S., Li, S., Liu, Q., Wu, K., Zhang, J., Wang, S., et al. (2015). The OsSPL16-GW7 regulatory module determines grain shape and simultaneously improves rice yield and grain quality. *Nat. Genet.* 47 (8), 949–954. doi: 10.1038/ng.3352

Wang, J., Liu, L., Guo, Y., Wang, Y., Zhang, L., Jin, L., et al. (2014). A dominant locus, qBSC-1, controls β subunit content of seed storage protein in soybean *(Glycine max* (L.) merri.). *J. Integr. Agr.* 13 (9), 1854–1864. doi: 10.1016/S2095-3119(13)60579-1

Warrington, C. V., Abdel-Haleem, H., Hyten, D. L., Cregan, P. B., Orf, J. H., Killam, A. S., et al. (2015). QTL for seed protein and amino acids in the benning × danbaekkong soybean population. *Theor. Appl. Genet.* 128 (5), 839–850. doi: 10.1007/s00122-015-2474-4

Weng, J., Gu, S., Wan, X., Gao, H., Guo, T., Su, N., et al. (2008). Isolation and initial characterization of GW5, a major QTL associated with rice grain width and weight. *Cell Res.* 18 (12), 1199–1209. doi: 10.1038/cr.2008.307

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 (1), 485. doi: 10.1186/s12870-020-02681-0

Xian, P., Cai, Z., Jiang, B., Xia, Q., Cheng, Y., Yang, Y., et al. (2022). GmRmd1 encodes a TIR-NBS-BSP protein and confers resistance to powdery mildew in soybean. *Plant Commun.* 3 (6), 100418. doi: 10.1016/j.xplc.2022.100418

Xu, Y., Li, H., Li, G., Wang, X., Cheng, L., and Zhang, Y. (2011). Mapping quantitative trait loci for seed size traits in soybean (*Glycine max* l. merr.). *Theor. Appl. Genet.* 122 (3), 581–594. doi: 10.1007/s00122-010-1471-x

Yang, K., Moon, J., Jeong, N., Chun, H., Kang, S., Back, K., et al. (2011). Novel major quantitative trait loci regulating the content of isoflavone in soybean seeds. *Genes Genom.* 33 (6), 685–692. doi: 10.1007/s13258-011-0043-z

Yang, H., Wang, W., He, Q., Xiang, S., Tian, D., Zhao, T., et al. (2017). Chromosome segment detection for seed size and shape traits using an improved population of wild soybean chromosome segment substitution lines. *Physiol. Mol. Biol. Plants.* 23 (4), 877–889. doi: 10.1007/s12298-017-0468-1

Yan, L., Li, Y., Yang, C., Ren, S., Chang, R., Zhang, M., et al. (2014). Identification and validation of an over-dominant QTL controlling soybean seed weight using populations derived from *Glycine max* × *Glycine soja*. *Plant Breeding.* 133 (5), 632–637. doi: 10.1111/pbr.12197

Yao, D., Liu, Z. Z., Zhang, J., Liu, S. Y., Qu, J., Guan, S. Y., et al. (2015). Analysis of quantitative trait loci for main plant traits in soybean. *Genet. Mol. Res.* 14 (2), 6101–6109. doi: 10.4238/2015.June.8.8

Yesudas, C. R., Bashir, R., Geisler, M. B., and Lightfoot, D. A. (2013). Identification of germplasm with stacked QTL underlying seed traits in an inbred soybean population from cultivars Essex and forrest. *Mol. Breeding.* 31 (3), 693–703. doi: 10.1007/s11032-012-9827-3

Yue, Y., Liu, N., Jiang, B., Li, M., Wang, H., Jiang, Z., et al. (2017). A single nucleotide deletion in j encoding GmELF3 confers long juvenility and is associated with adaption of tropic soybean. *Mol. Plant* 10 (4), 656–658. doi: 10.1016/j.molp.2016.12.004

Zeng, Z. B. (1994). Precision mapping of quantitative trait loci. *Genet. (Austin).* 136 (4), 1457–1468. doi: 10.1093/genetics/136.4.1457

Zhang, J., Song, Q., Cregan, P. B., and Jiang, G. (2016). Genome-wide association study, genomic prediction and marker-assisted selection for seed weight in soybean (*Glycine max*). *Theor. Appl. Genet.* 129 (1), 117–130. doi: 10.1007/s00122-015-2614-x

Zhang, X., Wang, J., Huang, J., Lan, H., Wang, C., Yin, C., et al. (2012). Rare allele of *OsPPKL1* associated with grain length causes extra-large grain and a significant yield increase in rice. *Proc. Natl. Acad. Sci.* 109 (52), 21534–21539. doi: 10.1073/pnas.1219776110

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 18 January 2023
DOI 10.3389/fpls.2022.1104022

OPEN ACCESS

EDITED BY
Zhengjun Xia,
Chinese Academy of Sciences (CAS), China

REVIEWED BY
Ying Huang,
UMR9213 Institut des Sciences des Plantes de Paris Saclay (IPS2), France
Yuefeng Guan,
Fujian Agriculture and Forestry University, China
Hai Nian,
South China Agricultural University, China

*CORRESPONDENCE
Chunyan Liu
cyliucn@126.com
Hongwei Jiang
j3994102@126.com
Qingshan Chen
qshchen@126.com

†These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION
This article was submitted to Functional and Applied Plant Genomics, a section of the journal Frontiers in Plant Science

RECEIVED 21 November 2022
ACCEPTED 23 December 2022
PUBLISHED 18 January 2023

CITATION
Cao F, Wei R, Xie J, Hou L, Kang C, Zhao T, Sun C, Yang M, Zhao Y, Li C, Wang N, Wu X, Liu C, Jiang H and Chen Q (2023) Fine mapping and candidate gene analysis of proportion of four-seed pods by soybean CSSLs. *Front. Plant Sci.* 13:1104022. doi: 10.3389/fpls.2022.1104022

Fine mapping and candidate gene analysis of proportion of four-seed pods by soybean CSSLs

Fubin Cao[1†], Ruru Wei[1†], Jianguo Xie[2†], Lilong Hou[1], Chaorui Kang[1], Tianyu Zhao[1], Chengcheng Sun[1], Mingliang Yang[1], Ying Zhao[1], Candong Li[3], Nannan Wang[3], Xiaoxia Wu[1], Chunyan Liu[1*], Hongwei Jiang[2*] and Qingshan Chen[1*]

[1]College of Agriculture, Key Laboratory of Soybean Biology in Chinese Ministry of Education, Northeast Agricultural University, Harbin, China, [2]Jilin Academy of Agricultural Sciences, Soybean Research Institute, Changchun, Jilin, China, [3]Jiamusi Branch Institute, Heilongjiang Academy of Agricultural Sciences, Jiamusi, Heilongjiang, China

Soybean yield, as one of the most important and consistent breeding goals, can be greatly affected by the proportion of four-seed pods (PoFSP). In this study, QTL mapping was performed by PoFSP data and BLUE (Best Linear Unbiased Estimator) value of the chromosome segment substitution line population (CSSLs) constructed previously by the laboratory from 2016 to 2018, and phenotype-based bulked segregant analysis (BSA) was performed using the plant lines with PoFSP extreme phenotype. Totally, 5 ICIM QTLs were repeatedly detected, and 6 BSA QTLs were identified in CSSLs. For QTL (*qPoFSP13-1*) repeated in ICIM and BSA results, the secondary segregation populations were constructed for fine mapping and the interval was reduced to 100Kb. The mapping results showed that the QTL had an additive effect of gain from wild parents. A total of 14 genes were annotated in the delimited interval by fine mapping. Sequence analysis showed that all 14 genes had genetic variation in promoter region or CDS region. The qRT –PCR results showed that a total of 5 candidate genes were differentially expressed between the plant lines having antagonistic extreme phenotype (High PoFSP > 35.92%, low PoFSP< 17.56%). The results of haplotype analysis showed that all five genes had two or more major haplotypes in the resource population. Significant analysis of phenotypic differences between major haplotypes showed all five candidate genes had haplotype differences. And the genotypes of the major haplotypes with relatively high PoFSP of each gene were similar to those of wild soybean. The results of this study were of great significance to the study of candidate genes affecting soybean PoFSP, and provided a basis for the study of molecular marker-assisted selection (MAS) breeding and four-seed pods domestication.

KEYWORDS

wild soybean, CSSLs, BSA, secondary segregation population, QTL for the proportion of four-seed pods, candidate gene

1 Introduction

Soybean originated in China and then spread around the world. As an important food crop and oil crop, it has a cultivation history of more than 5000 years (Guo, 2004; Barik et al., 2018; Zhang et al., 2018). China is the world's largest soybean consumer. In the past 10 years, China's soybean consumption accounted for more than 30% of the world's total, ranking first in the world (Liu and Fan, 2021). However, China's soybean yield per unit is low, only 70% of the international average (Si and Han, 2021), and the domestic soybean supply is weak, with a supply rate of only approximately 15% (Liu and Fan, 2021). Improving soybean yield is an urgent problem to be solved in the world's soybean industry.

The yield of soybean is affected by multiple factors, such as 100-seed weight, the number of pods per plant, the number of seeds per plant and the number of seeds per pod. These are all quantitative traits controlled by multiple genes and are easily affected by environmental factors (Xie et al., 2021). The total number of soybean pods per plant is one of the main factors limiting soybean yield. When the total number of pods varies little or the potential for increasing the number of pods is limited, PoFSP becomes the main limiting factor for the seed number per plant of soybean. Studies have shown that varieties with a high PoFSP have higher yield and productivity, and it is an effective way to increase the yield by increasing PoFSP of varieties, which has great practical value for increasing the yield. (Peng et al., 1994).

Compared with traditional breeding, the application of molecular marker technology has mostly overcome the difficulties introduced by environmental factors into phenotype identification (Song et al., 2012), and the emergence of MAS and QTL mapping has significantly shortened the breeding years (Hasan et al., 2021). MAS indirectly realizes the selection of QTL controlling a certain trait through the selection of genetic markers, so as to achieve the purpose of selecting the trait (Wakchaure et al., 2015). In 1988, Paterson mapped QTLs by using molecular markers in tomato for the first time (Paterson et al., 1988), and this method has been widely used in different crops (Song et al., 2005; Jiang et al., 2011; Yao et al., 2015; Botero-Ramírez et al., 2020). At present, a large number of studies have been carried out on QTL mapping of pod traits in soybean, and many QTLs related to four-seed pods have been mapped. Wang et al. mapped the QTL related to the number of pods and compared the heritability of one, two, three, and four seed pods by using soybean RILs (recombinant inbred line), and they found that the heritability of four-seed pods was higher and the main reason for the variation in seed pods number was the shedding of flower pods. In 2007, Wang et al. mapped two four-seed pods QTLs detected in two years by using CIM (composite interval mapping) method in soybean RILs (Wang et al., 2007). In 2008, Li et al. mapped three four-seed pods QTLs by using CIM method in soybean CSSLs (Li et al., 2008). In 2009, Zhou et al. mapped two four-seed pods QTLs by using CIM method in soybean RILs (Zhuo et al., 2009). In 2012, Gao et al. mapped 8 QTLs associated with four-seed pods by using CIM method in soybean RILs (Gao et al., 2012). In 2013, Yang et al. mapped 4 four-seed pods QTLs by using the multi-environment joint analysis method in soybean RILs (Yang et al., 2013b). In 2018, Ning et al. mapped 36 and 12 four-seed pods QTLs in soybean by using single marker analysis, CIM and multiple interval mapping methods RIL6013 and RIL3613, respectively (Ning et al., 2018). In 2021, Li et al. constructed 3 RHL s and used single marker analysis to fine-map a four-seed pods QTL, narrowing the QTL interval to 321 kb (Li et al., 2021).

In 1991, BSA method was first proposed by Michelmore et al. (Michelmore et al., 1991). Its core purpose is to use extreme material of the target trait in the population to construct two mixed pools and to locate the target gene by analyzing the degree of association between the polymorphic molecular markers and the phenotype (Zou et al., 2016). Compared to using near-isogenic lines (NILs) to identify markers in specific regions of the genome, BSA is faster and less labor. And BSA significantly reduces the cost of sequencing and analysis because only a few extreme individuals in a population are needed (Yang et al., 2013a).

The quality of the genetic population directly affects the effect, difficulty and application scope of the constructed genetic map s. If the genome differences between the parents are greater, the effect of exchange and recombination between their chromosomes will be more obvious, and the DNA polymorphism and phenotypic variation will be more abundant (Xi et al., 2005). Cultivated soybean was domesticated from annual wild soybean (Broich and Palmer, 1980), and there are significant differences between them. In this study, CSSLs were constructed using cultivated soybeans and wild soybeans. There are abundant variations in this population, and the ability to detect QTLs is strong. Although the genetic variation information is rich, the diversity of their genetic background will affect the accuracy of QTL mapping by using traditional mapping population such as F_2, BC, DH, RIL and others (Zhao et al., 2009; Qiao et al., 2016). Compared with the primary mapping population, the genetic background of CSSLs is purer, which not only eliminates genetic noise but also improves the accuracy of QTL mapping (Li, 2014). At present, CSSLs have been applied to 20 crops, and researchers have used

Abbreviations: CSSLs, Chromosome segment substitution lines; BSA, Bulked Segregant Analysis; RHL, Residual Heterozygous Lines; BLUE, Best Linear Unbiased Estimator; MAS, Marker-assisted selection; SNP, Single Nucleotide Polymorphisms; InDel, Insertion-Deletion; ANOVA, Analysis of Variance; ICIM, Inclusive Composite Interval Mapping; CIM, Composite Interval Mapping; PoFSP, The proportion of four-seed pods; QTL, Quantitative trait loci; LOD, Logarithm of odds; PVE, Phenotypic Variance Explained; qRT-PCR, Quantitative Real-time PCR; RILs, Recombinant Inbred Line.

these CSSLs to map a large number of QTLs (He et al., 2014; Li et al., 2016; Alyr et al., 2020; Yuan et al., 2020; Xu et al., 2022).

In this study, CSSLs constructed previously in the laboratory was used as the material. Combined with the results of ICIM method and BSA method in CSSLs, PoFSP QTL in soybean was initially mapped, and the secondary segregated populations were constructed for fine mapping. The genes were screened out in the fine mapping interval, and the candidate genes for PoFSP were analyzed and predicted by expression and haplotype analysis.

2 Materials and methods

2.1 Plant material and field management

The population orientation material for this study was CSSLs containing 208 lines constructed from SN14 (the recurrent parent) and ZYD00006 (Jiang et al., 2020), which were planted in Xiangyang Farm, Harbin, China, from 2016 to 2018 (45°45′N, 126°38′E). A completely random block design was adopted, and each line was planted in one row, repeated three times. The row length was 5 m, the row spacing was 65 cm, and the plant spacing in each single-row plot was 6 cm, with approximately 80 plants. Fine mapping population was constructed by backcrossing R92 with extremely high PoFSP in CSSLs with recurrent parent SN14. The R92-F_1 seeds obtained in the same year were propagated in Nannan, Hainan, and the R92-F_2 was planted in Xiangyang Farm the following year, and 121 R92-F_2 individuals were obtained. RHLs(H1) was constructed from an individual in R92-F_2 whose genotype was completely heterozygous in the target interval and whose remaining background was relatively pure. The resource population was also planted on the Sunshine Farm, and the planting method was the same as that of CSSLs. The field management of the above materials followed general agricultural practices.

2.2 Phenotypic identification and statistical analysis

For CSSLs, 5 complete plants were selected for each material at the mature stage in the field to test the number of pods and the number of seeds per pod at the end of September every year before harvesting. Each plant of R92-F_2 and H1 was also investigated in the field before harvesting for the number of pods and the number of seeds per pod, and PoFSP was calculated. The measurement standard referred to the "Soybean Germplasm Description Specification and Data Standard" (Zhou et al., 2015).

The phenotypic data were sorted using Microsoft® Excel 2016, the descriptive analysis was performed using SPSS 17.0, and the significance analysis was performed using one-way ANOVA. BLUE was calculated by the BLUE calculation option in AOV module of ICIMapping 4.2 software. The parameters used were each line of CSSLs as a fixed factor and the years and repetition as random factors.

2.3 Bulk segregant analysis

In this study, phenotype-based BSA was performed by 30 materials with high PoFSP (PoFSP > 17.14%) and 30 materials with low PoFSP (PoFSP< 1.46%) in 208 CSSLs. A total of 3,716,818 SNPs were detected, and 3,105,246 high-quality SNPs were obtained after screening out. The Euclidean Distance method was used to analyze the association between the markers and PoFSP in the sequencing results:

ED=

$$\sqrt{(A1-A2)^2+(C1-C2)^2+(G1-G2)^2+(T1-T2)^2}$$

A1 and A2 was in the same position. A1 represented the occurrence frequency of A in the high phenotype pool, and A2 represented the occurrence frequency of A in the low phenotype pool. C, G, and T were the same as A. The mean ± 3 times the standard deviation (0.02238) was taken as the threshold value to judge whether the marker and the trait were closely linked, and if it exceeded the specified threshold, it was considered to be related to the trait.

2.4 QTL mapping for PoFSP

The whole genome resequencing data of CSSLs (Zheng et al., 2022) constructed in the laboratory and PoFSP data and BLUE values from 2016-2018 were used for QTL mapping. QTL mapping employed the RSTEP-LRT-ADD model with the 1,000 permutations calculation of ICIMapping 4.2 software. Parameters setting was "By condition number" = −1,000 (equivalent to deleting duplicate markers), "PIN" = 0.001 (PIN: the largest-value for entering variables in stepwise regression of residual phenotype on marker variables). DNA from fresh leaves was extracted using the cetyltrimethylammonium ammonium bromide (CTAB) plant tissue DNA extraction method, and the DNA concentration and purity were determined using a NanoDrop 2000C (Sunnyvale, California, USA) ultradifferential photometer and 1.5% agarose gel electrophoresis. The SSR markers were used as molecular markers. After the polymerase chain reaction (PCR) of Panaud, the genotype of the secondary segregated population was detected by 8% PAGE separation (Doyle, 1990). The genetic maps of the secondary segregating populations were constructed by the MAP module in ICIMapping 4.2 software. QTL mapping employed ICIM-ADD model with 1,000 permutations calculation of ICIMapping 4.2 software. And the threshold of ICIMapping for QTL detection was chose $p \leq 0.05$.

2.5 qRT–PCR analysis of candidate genes

According to different stages of flower bud differentiation, the flower tissues of soybean were sampled three days before flowering, two days before, one day before, on the day of flowering, three days after flowering and five days after flowering and stored in liquid nitrogen. The sampling materials were selected from the R92-F_2. L-9 and L-14 which had the genotype of ZYD00006 had a high PoFSP (PoFSP > 35.92%). L-72 and L-93 which had the genotype of SN14 had a low PoFSP (PoFSP< 17.56%). Three biological replicates were sampled per material. After grinding the samples in liquid nitrogen, RNA from the plant tissues was extracted using the TRIzol method. The RNA concentration and purity were determined using a NanoDrop 2000C (Sunnyvale, California, USA) ultradifferential photometer and 1.5% agarose gel electrophoresis. Reverse transcription of the extracted RNA into cDNA using the Tianhe Real-time quantitative PCR (RT–qPCR) kit was performed using SYBR qPCR Mix (Vazyme, Q711, Vazyme Biotech, Nanjing, China) on the Light Cycler 480 System (Roche, Roche Diagnostics, Basel, Switzerland). qRT–PCR primer sequences for the candidate genes were designed using Premier 5.0. Using *GmACTIN* as an internal reference, the average of three biological replicates was taken, and the relative expression levels of the candidate genes were calculated by the $2^{-\Delta Ct}$ method.

2.6 Haplotype analysis

Haplotype analysis of PoFSP candidate genes was conducted using 527 soybean germplasm resources. Combined with the sequence information of the promoter region and CDS region of the candidate gene in the reference genome, local sequence alignment was performed on the resequencing data of the candidate gene between the parents of CSSLs to find the difference sites and analyze the variation in the promoter elements and CDS region translation products. The haplotype distribution of the candidate genes in the resource population was counted by DnaSP 5.0 software. The major haplotypes were divided by the number of varieties exceeding 5% of in resource population, and Haploview 4.2 software was used to perform linkage disequilibrium (LD) analysis on the mutation sites among the major haplotypes. Using one-way AOV(analysis of variance) in IBM SPSS Statistics 22 software, combined with the phenotypic data of PoFSP in the resource population in 2021, the significance of the difference between the phenotypes of each major haplotype material was compared.

3 Results

3.1 QTL analysis in CSSLs

QTL analysis was performed based on the phenotypic data (Zheng et al., 2022) and BLUE value data of PoFSP in CSSLs from 2016 to 2018 (Figure 1). A total of 17 QTLs were detected, distributed on 13 chromosomes (Supplementary Table 1), of which 5 were repeatedly detected in two or more years. *qPoFSP07-1* and *qPoFSP07-2* on Chr07 were detected in 2017 and BLUE, with maximum PVE (phenotypic variance explained) of 3.87% and 8.76%, respectively; *qPoFSP13-1* on Chr13 was detected in 2018 and BLUE, with a maximum PVE of 9.18%; *qPoFSP17-1* on Chr17 was detected in 2018 and BLUE, with a maximum PVE of 5.14%; *qPoFSP20-2* on Chr20 was detected in 2017, 2018 and BLUE, with a maximum PVE of 30.69% (Table 1; Figure 2). The above 5 QTLs were repeatedly detected in two or more years, which was more reliable than the QTLs that appeared in a single year. And the five QTLs would be the focus of follow-up research.

3.2 Analysis of the BSA sequencing results

BSA-seq was performed on the extreme materials of PoFSP in CSSLs. A total of 6 PoFSP QTLs were detected, which were distributed on Chr01, Chr03, Chr08, Chr12, Chr13 and Chr18, and the interval size was 0.09 Mb, 3.67 Mb, 0.10 Mb, 2.57 Mb, 3.61 Mb, and 27.4 Mb, respectively (Supplementary Table 2; Figure 3A). The QTL located on chromosome 13 coincides with the *qPoFSP13-1* in CSSLs mapping results, with a size of 3.61 Mb (Figure 3B). *qPoFSP13-1* was repeated twice in ICIM results and overlapped with BSA result s, showing a strong association with PoFSP phenotype. Therefore, *qPoFSP13-1* was further fine-mapped as an important candidate QTL for PoFSP.

3.3 Construction and fine mapping of secondary separation populations

According to the location information of the candidate QTL (*qPoFSP13-1*) on the chromosome and the whole genome resequencing information of CSSLs, combined with the introduction of the ZYD00006 segment in the whole genome, a total of 16 lines from 208 CSSLs were screened out for the ZYD00006 homozygous genotype introduction segment completely covered the candidate QTL on chromosome 13. The results of phenotypic identification showed that PoFSP in Line R92 was significantly higher than that of its parents and reached an extremely significant level ($P \leq 0.01$), which was a superparental

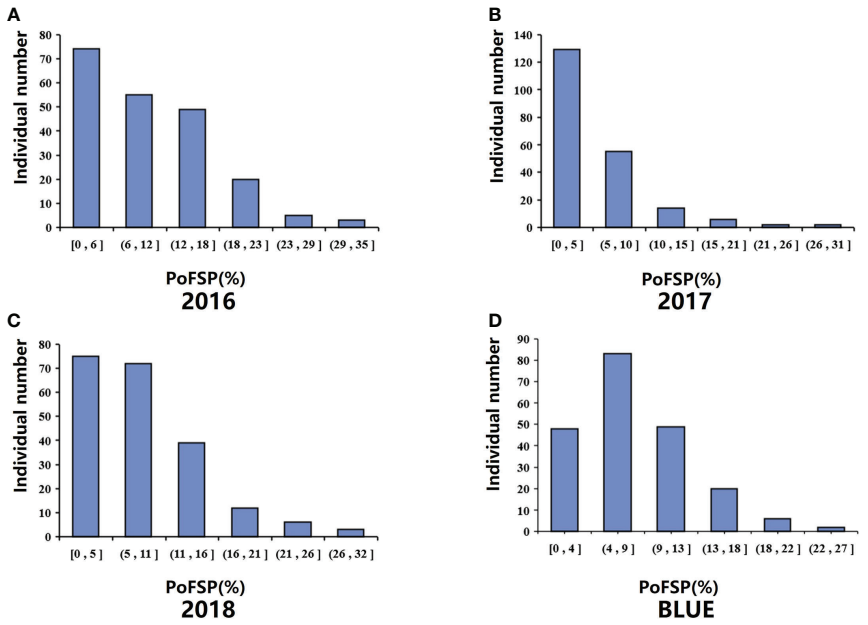


FIGURE 1
Phenotypic identification of PoFSP in CSSLs. **(A)**: PoFSP of CSSLs in 2016. **(B)**: PoFSP of CSSLs in 2017. **(C)**: PoFSP of CSSLs in 2018. **(D)**: PoFSP of CSSLs in BLUE.

material, and its target introduction segment size was 4.9 Mb (Figures 4A, B). Therefore, R92 was selected as the follow-up research material.

The secondary segregation population R92-$F_{2,}$ including 121 plants, was constructed by crossing R92 with the recurrent parent. PoFSP was calculated for the R92-F_2. The phenotypic data showed that the highest and lowest PoFSP in R92-F_2 were 56.00% and 2.38%, respectively. The R92-F_2 was rich in phenotypic variation with obvious differences and was suitable for QTL analysis (Figure 5A; Table 2). Eight pairs of SSR markers evenly distributed on the R92 target introduction segment and polymorphic between parents were screened out, linkage analysis was performed on the R92-F_2, and the QTL interval was located between markers 13-755 and 13-769. The length of the candidate interval was 1.2 Mb, and PVE was 11.06%. The ADD results showed that the additive effect of the gain came from ZYD00006 (Figure 5B; Table 3).

RHLs(H1) is constructed from an individual in R92-F2 whose genotype is completely heterozygous in the target

TABLE 1 PoFSP QTLs detected in CSSLs for two or more years.

Years	Name	Chromosome	LOD	PVE (%)	ADD	Start (bp)	End (bp)
2017	*qPoFSP07-1*	Chr07	3.46	3.87	0.03	3900022	3934584
2017	*qPoFSP07-2*	Chr07	7.19	8.76	0.02	39710561	39776712
2017	*qPoFSP20-2*	Chr20	6.94	8.35	-0.02	35907647	35932855
2018	*qPoFSP13-1*	Chr13	4.22	6.69	0.02	23976621	24036668
2018	*qPoFSP17-1*	Chr17	3.32	5.14	-0.03	2678066	2858489
2018	*qPoFSP20-2*	Chr20	9.13	15.12	-0.04	35907647	35932855
BLUE	*qPoFSP07-1*	Chr07	2.96	1.94	0.02	3900022	3934584
BLUE	*qPoFSP07-2*	Chr07	10.50	7.44	0.02	39710561	39776712
BLUE	*qPoFSP13-1*	Chr13	12.63	9.18	0.02	23976621	24036668
BLUE	*qPoFSP17-1*	Chr17	4.65	3.05	-0.02	2678066	2858489
BLUE	*qPoFSP20-2*	Chr20	33.31	30.69	-0.05	35907647	35932855

the "+" additive indicates that the additive effect comes from the allele of the wild parent ZYD00006.

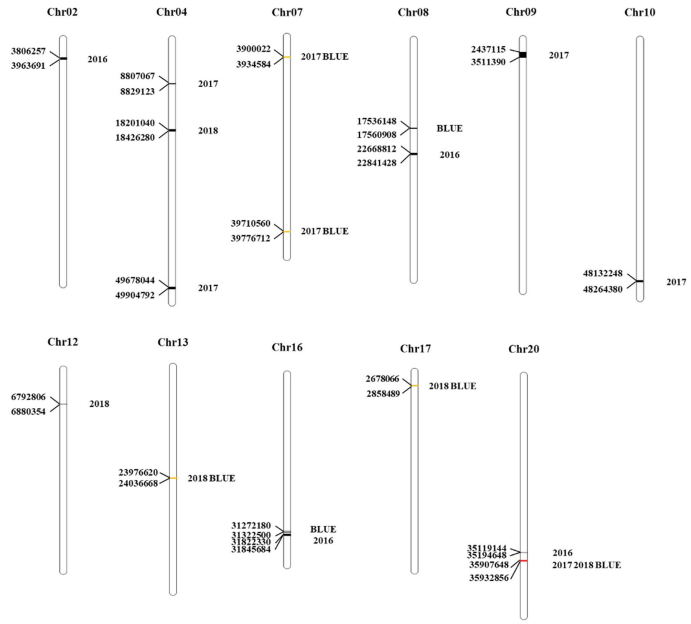


FIGURE 2
Distribution of PoFSP QTLs on chromosomes in CSSLs. The unit of each position in the figure is bp.

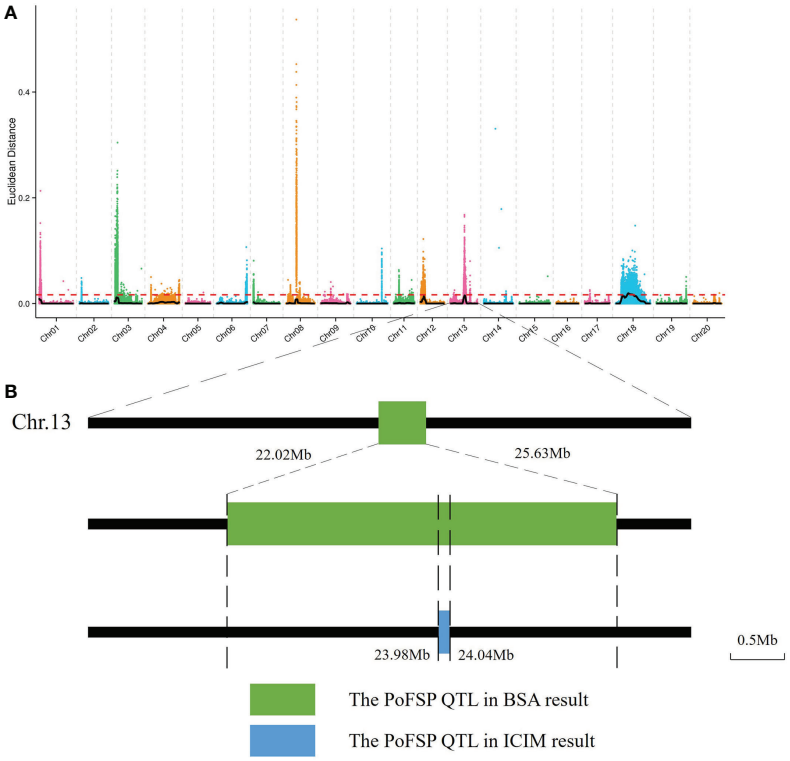


FIGURE 3
BSA results. **(A)**: The distribution of Euclidean distance (ED)-associated values on chromosomes. **(B)**: Preliminary mapping process of QTLs for PoFSP. [a]: The abscissa is the chromosome name. The color points represent the ED value of each single nucleotide polymorphism (SNP) locus. The black line is the fitted ED value, and the red dotted line represents the significantly associated threshold. The higher the ED value is, the better the correlation.

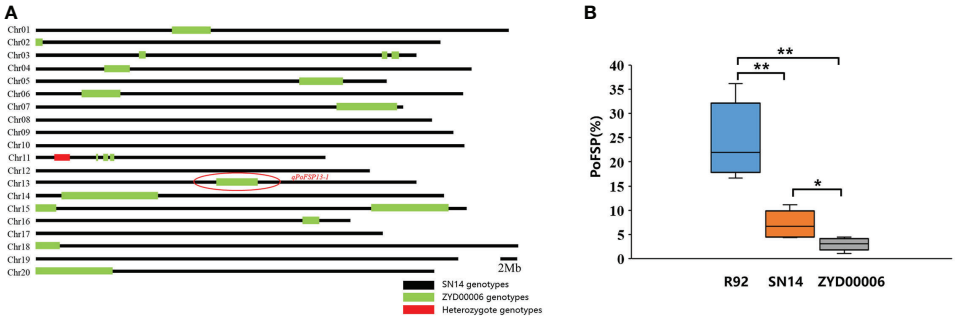


FIGURE 4
Genotypic background and phenotypic identification of R92. **(A)**: The distribution of the ZYD00006 segment in the R92 genome. **(B)**: PoFSP in R92 and the parents. [b]: "*" means the significant difference at the 0.05 level, and "**" means the significant difference at the 0.01 level.

interval and whose remaining background is relatively pure. In H1 containing 126 plants, the highest PoFSP was 47.22%, and the lowest PoFSP was 0.00. The H1 was rich in phenotypic variation with obvious differences and was suitable for QTL analysis (Figure 5C; Table 2). Ten pairs of SSR markers evenly distributed in the initial mapping QTL interval and polymorphic between parents were screened out, and H1 was genotyped. A total of one PoFSP QTL with PVE of 19.47% was detected. The ADD results showed that the additive effect of gain still came from ZYD00006 (Figure 5D; Table 3). Finally, *qPoFSP13-1* was fine-mapped between the two flanking markers 13-766 and 13-769, and the candidate interval length was 100 kb (Figure 6).

3.4 Gene screening and prediction within the QTL interval

According to the Williams 82.a2.v1 reference genome sequence information in SoyBase(https://www.soybase.org/) and Phytozome(https://phytozome-next.jgi.doe.gov/), a total of 14 PoFSP candidate genes were annotated in *qPoFSP13-1* (Supplementary Table 3). Combined with the resequencing data of the parents of CSSLs, the sequence analysis results showed 13 candidate genes had InDel mutations and 14 candidate genes had SNP mutations in the promoter region, 8 candidate genes had nonsynonymous mutations and 2 candidate genes had InDel mutations in the coding region (Supplementary

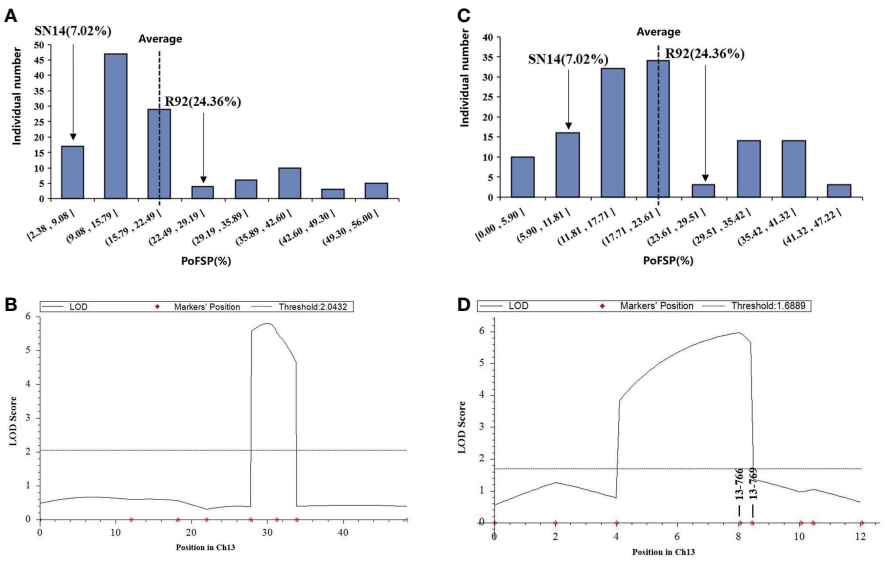


FIGURE 5
Phenotypic identification and QTL results of fine-mapping populations. **(A)**: Frequency histogram of the phenotypic distribution of PoFSP in the R92-F_2. **(B)**: QTL mapping results of ICIM method in R92-F_2. **(C)**: Frequency histogram of the phenotypic distribution of PoFSP in the H1. **(D)**: QTL mapping results of ICIM method in H1.

TABLE 2 Phenotypic statistics of PoFSP in the secondary segregated population.

Population	Number	Average PoFSPN (%)	Minimum (%)	Maximum (%)	PoFSPN for SN14 (%)	PoFSPN for R92 (%)
R92-F_2	121	19.16	2.38	56.00	7.02	24.36
H1	126	20.22	0.00	47.22		

Table 4; Supplementary Table 5; Supplementary Table 6). The expression levels of 14 candidate genes were analyzed in 6 periods before and after flowering. qRT-PCR results showed that *Glyma.13G126000* was expressed at low levels before flowering, on the day of flowering and after flowering in the high-extreme materials L-9 and L-14. However, its expression level gradually increased before flowering, peaked on the day of flowering, and then gradually decreased in the low-extreme materials L-72 and L-93. *Glyma.13G126100* was also expressed at low levels during each period of in the high-extreme materials L-9 and L-14. However, its expression in low-extreme materials L-72 and L-93 was higher than that in high-extreme materials as a whole, and its expression also increased gradually before flowering and peaked on the day of flowering, and then gradually decreased in L-72 and L-93. The expression levels of *Glyma.13G126300, Glyma.13G126400* and *Glyma.13G126700* gradually increased before flowering, peaked on the day of flowering, and then gradually decreased in the high-extreme materials L-9 and L-14. However, it showed low expression in each period in the low-extreme materials L-72 and L-93 (Figure 7). According to the above results, it is predicted that *Glyma.13G126000* and *Glyma.13G126100* might be negative regulators of PoFSP, while *Glyma.13G126300*, *Glyma.13G126400* and *Glyma.13G126700* might be positive regulators of PoFSP. In conclusion, *Glyma.13G126000*, *Glyma.13G126100*, *Glyma.13G126300*, *Glyma.13G126400*, and *Glyma.13G126700* were screened out as candidate genes affecting PoFSP in soybean.

3.5 Candidate gene haplotype analysis

Using DnaSP 5.0 software, haplotype analysis was performed on the above five candidate genes for PoFSP in the resource population contained 527 varieties. AOV was used to analyze the significance of the phenotypic differences among the major haplotypes. The major haplotypes were divided according to the standard that the number of varieties exceeded 5% of the resource population. All five candidate genes had two or more major haplotypes in the resource population.

Glyma.13G126000 had 2 major haplotypes in the resource population, Hap_2 had 367 resource varieties, and Hap_6 has 84 resource varieties. The results of AOV showed that there was an extremely significant difference in PoFSP between Hap_2 and Hap_6. Through SNP analysis, *Glyma.13G126000* had a total of 8 differential SNPs in the resource population, of which 7 SNPs were between Hap_2 and Hap_6. SNP-1009 was located on a motif sequence in the promoter region of *Glyma.13G126000*, and its mutation might affect the binding of transcription factors to promoters. SNP-1270 mutation did not cause changes in promoter elements, but it was tightly linked to SNP-1009 according to the LD analysis. SNP-2667, SNP-2019, SNP-1560 and SNP-806 were also located in the promoter region, but none of their mutations caused changes in promoter elements. In the coding region, SNP3498 was A in Hap_2, which was consistent with ZYD00006, while it was G in Hap_6, which was consistent with SN14 (Figures 8A, B). This site mutation caused the corresponding translation product to change from Glu to Arg, which might affect the changes in gene structure or function.

Glyma.13G126100 had a total of 2 major haplotypes in the resource population, Hap_2 had 114 resource varieties, and Hap_3 had 362 resource varieties. The results of AOV showed that there was an extremely significant difference in PoFSP between Hap_2 and Hap_3. Through SNP analysis, *Glyma.13G126100* had a total of 6 differential SNPs in the resource population. SNP-1991 was located on the promoter of *Glyma.13G126100*, and its mutation resulted in an alteration of an MYB transcription factor-related element that might affect gene transcription. SNP-1907 did not cause changes in promoter elements, but it was tightly linked to SNP-1991 by LD analysis. SNP-2773, SNP-2719, and SNP-1227 were also located in the promoter region, but none of their mutations caused changes in the promoter elements. In the coding region, SNP23 was C in Hap_2, which was consistent with SN14 and encoded Pro, while

TABLE 3 Mapping of PoFSP QTLs by secondary segregated population ICIM.

Population	Chromosome	Left marker	Right marker	LOD	PVE (%)	Add	Size (Mb)
R92-F_2	Chr13	13-755	13-769	2.78	11.06	0.06	1.2
H1	Chr13	13-766	13-769	6.00	19.47	0.05	0.1
the "+" additive indicates that the additive effect comes from the allele of the wild parent ZYD00006.							

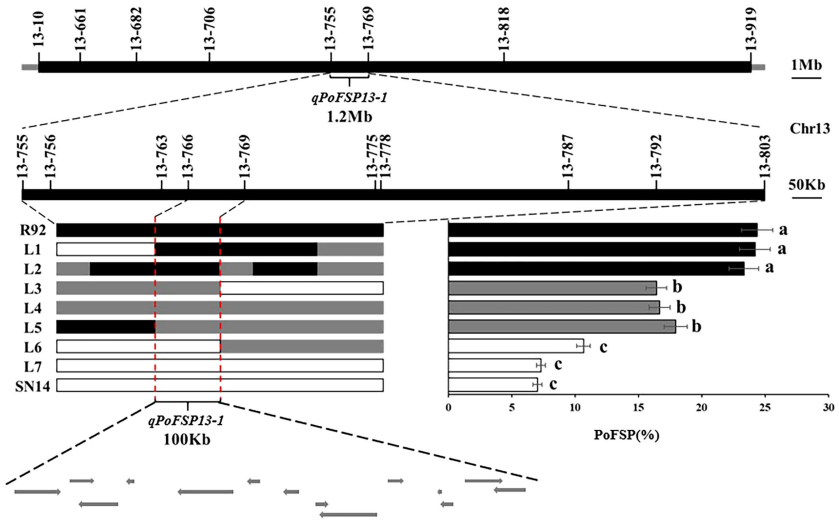


FIGURE 6

Schematic diagram fine mapping of *qPoFSP13-1*. ■ represents the ZYD00006 genotype, □ represents the SN14 genotype, and ▒ represents the heterozygous genotype. R92, L1 and L2 were the same genotypes as ZYD00006; L3, L4 and L5 were heterozygous genotypes of ZYD00006, and L6 and L7 were the same genotypes as SN14. Different letters indicate the significant differences, while the same letters indicate the non-significant differences.

it was G in Hap_3, which was a dominant mutation type different from its parents and encoded Gly (Figures 8C, D). Variations in the translation products caused by the mutations at this site might affect the gene structure or function.

Glyma.13G126300 had a total of 2 major haplotypes in the resource population, Hap_2 had 78 resource varieties, and Hap_3 had 364 resource varieties. The results of AOV showed that there was an extremely significant difference in PoFSP between Hap_2 and Hap_3. Through SNP analysis, *Glyma.13G126300* had a total of 15 differential SNPs in the resource population, of which 5 SNPs were between Hap_2 and Hap_3, and all of them were located in the promoter region. There was a motif sequence on SNP-443 and SNP-284 respectively, and thier mutations caused changes in the motif sequence, which might affect transcription factor binding. According to LD analysis, SNP-592 was closely linked with SNP-443 and SNP-284, but SNP-592 did not cause the promoter element to change (Figures 9A, B).

Glyma.13G126400 has a total of 2 major haplotypes in the resource population, Hap_1 has 128 resource varieties, and Hap_3 has 364 resource varieties. The results of AOV showed that there was an extremely significant difference in PoFSP between Hap_1 and Hap_3. Through SNP analysis, *Glyma.13G126400* has 5 differential SNPs in the resource population, of which 3 SNPs were between Hap_1 and Hap_3, and all of them are located in the promoter region. According to LD analysis, SNP-1785 and SNP-1499 are closely linked with SNP-1723, but mutation of them will not cause changes in the promoter elements, and there is a TATA-box sequence at SNP-1723. Mutations in this sequence may affect the production of TATA-box-binding proteins, possibly resulting in the inhibition of RNA polymerase transcription factor synthesis (Figures 10A, B).

Glyma.13G126700 had a total of 4 major haplotypes in the resource population. Hap_1 had 75 resource varieties, Hap_3 had 361 resource varieties, Hap_4 had 28 resource varieties, and Hap_6 had 28 resource varieties. AOV and multiple comparisons showed that only Hap_3 and Hap_4 had a significant difference in PoFSP. Through SNP analysis, *Glyma.13G126700* had a total of 5 SNPs in the resource population, of which 2 SNPs were between Hap_3 and Hap_4. SNP-2213 was located in a CAAT-box sequence on the promoter, and its mutation might lead to changes in the CAAT-box sequence, which might affect the normal initiation of gene transcription. SNP1410 was located in the 3'UTR, and its mutation might affect the stability of the gene structure. According to LD analysis, although SNP1571 had no mutation in Hap_3 and Hap_4, it was closely linked with SNP1410. In Hap_4, the linkage between the two was broken, which might also destroy the 3'UTR structure and affect the stability of the gene structure (Figures 10C, D).

4 Discussion

Soybean originated in China and is an important oil and cash crop (He et al., 2022). The yield of soybean is affected by a variety of yield factors, such as the number of pods per plant, the

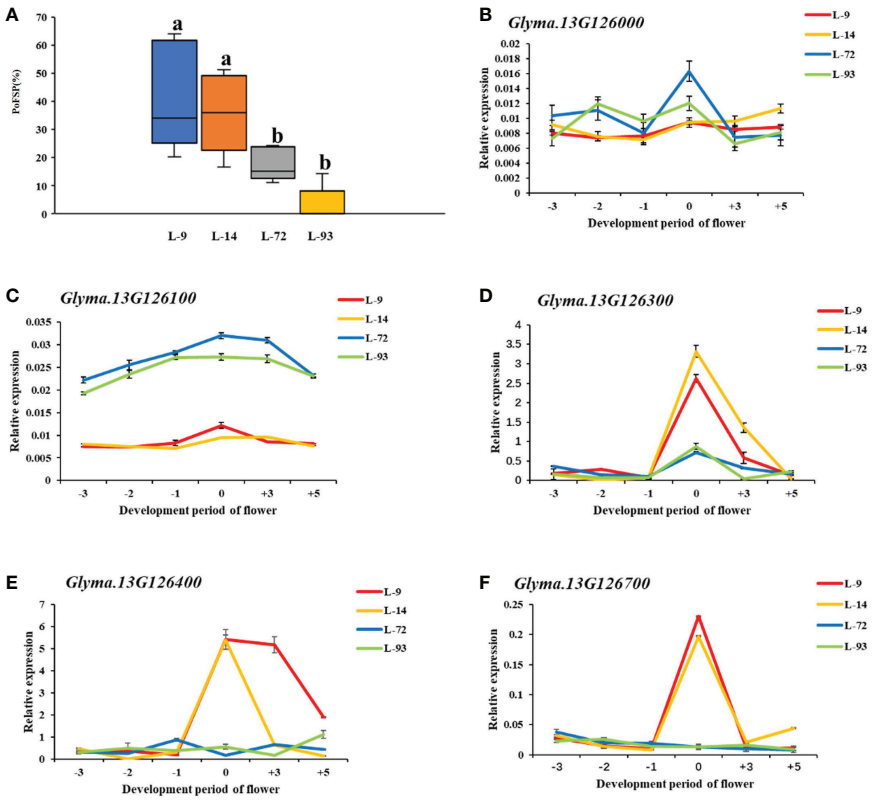


FIGURE 7
Expression analysis of candidate genes. **(A)**: Phenotypic identification of extreme materials. **(B–F)**: Real-time PCR results of the candidate genes. [a]: Different letters indicate the significant differences at the 0.05 level, while the same letters indicate the nonsignificant differences. [b-f]: -3, -2, -1, 0, + 3, +5 represent 3 days before flowering, 2 days before flowering, 1 day before flowering, the day of flowering, 3 days after flowering, 5 days after flowering, respectively.

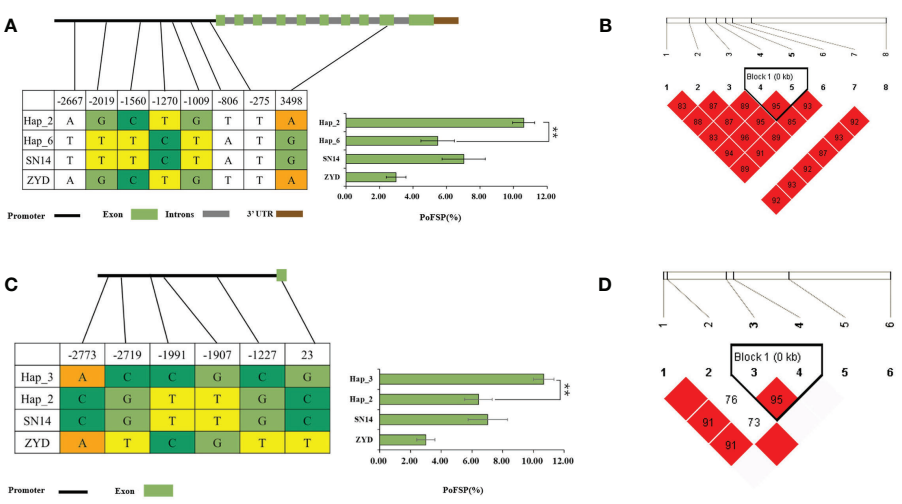


FIGURE 8
Haplotype analysis results of *Glyma.13G126000* and *Glyma.13G126100*.**(A)**: Correlation analysis of the major haplotype phenotype in *Glyma.13G126000*. **(B)**: LD analysis of SNPs among major haplotypes of*Glyma.13G126000*. **(C)**: Correlation analysis of the major haplotype phenotype in *Glyma.13G126100*. **(D)**: LD analysis of SNPs among major haplotypes of*Glyma.13G126100*. [a,c]: "**" indicates a significant difference at the 0.01 level.

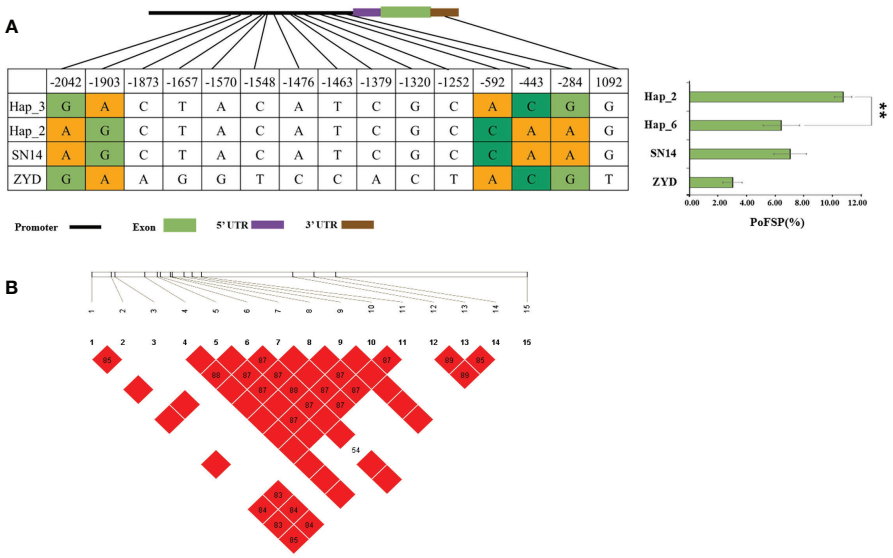


FIGURE 9
Haplotype analysis results of *Glyma.13G126300*.**(A)**: Correlation analysis of the major haplotype phenotype in *Glyma.13G126300*.**(B)**: LD analysis of SNPs among major haplotypes of*Glyma.13G126300*. [a]: "**" indicates a significant difference at the 0.01 level.

number of seeds per pod, 100-seed weight and the number of seeds per plant. It is a quantitative trait susceptible to environmental factors and controlled by multiple genes (Sun et al., 2012). As one of the important factors affecting yield, it is of great significance to improve PoFSP for yield research.

According to the data published on the SoyBase, 425 QTLs related to soybean four-seed pods have been reported to date. In this study, 12 QTLs detected in a single year and 5 QTLs repeatedly detected in two years or more were mapped by PoFSP data and BLUE value in CSSLs from 2016 to 2018. Compared with previous research results, *qPoFSP20-2* detected in this study in 2017, 2018 and BLUE value highly overlapped with the *Ln* loci mapped by Fang et al. (Fang et al., 2013) in 2013, reflecting the reliability of CSSLs mapping results. *qPoFSP20-2* was detected three times and its LOD and PVE were the most prominent. But *qPoFSP13-1* was detected two times and its LOD and PVE were also outstanding, and coincided with BSA results. In this study, the latter QTL was selected as the research object more convincingly.

In previous studies, the *Ln* gene named after a narrow leaflet has been frequently studied, and it was an important gene affecting the number of four-seed pods (Takahashi, 1934; Domingo, 1945; Peng et al., 1994). A large number of studies have proven that soybean leaf type and four-seed pods were controlled by pleiotropic genes (Johnson and Bernard, 1962; Weiss, 1970; Dinkins et al., 2002). Fang et al. successfully cloned the *Ln* gene and found through functional analysis that the *Ln* gene could increase the number of four-seed pods (Fang et al., 2013). Moreover, while the *Ln* gene has been studied in depth, other genes related to PoFSP have also been discovered one after another. Wang et al. successfully cloned the soybean *GmCYP78A10* gene in 2015 and found that *GmCYP78A10* can increase PoFSP (Wang et al., 2015). Wang et al. analyzed transcriptome sequencing to obtain differentially expressed genes in 2022 and found that the *Glyma.19G240800* gene was homologous to the Arabidopsis cytochrome P450 gene *Klu* (Wang et al., 2022). Poretska et al. found that the *Klu* gene was required for female meiosis during ovule development (Poretska et al., 2020). In this study, CSSLs and constructed RHLs were used to fine-map a QTL for PoFSP (*qPoFSP13-1*). In addition, *qPoFSP13-1* has not been reported in previous studies, and no previously reported PoFSP genes were found in *qPoFSP13-1*, including the genes mentioned above. Therefore, this study might predict a new gene affecting PoFSP in soybean.

China's wild soybean resources are rich and widely distributed. Compared with cultivated soybeans, wild soybeans have more pods and a higher protein content (Jin et al., 2017). The application of wild soybean varieties can effectively solve the problem of genetic simplification in the breeding of cultivated soybean. This study used PoFSP phenotype data and BLUE values from 2016-2018 of the wild soybean CSSLs and BSA results to comap the QTL(*qPoFSP13-1*), and the genes identified by fine mapping with high reliability. The mapping results showed that the QTL had an additive effect of gain from the wild parent. Moreover, the genotypes of the five candidate genes in the resource population with higher haplotypes of PoFSP were similar to those of wild soybean. This indicated that the original gene haplotype affecting PoFSP in wild soybean has been

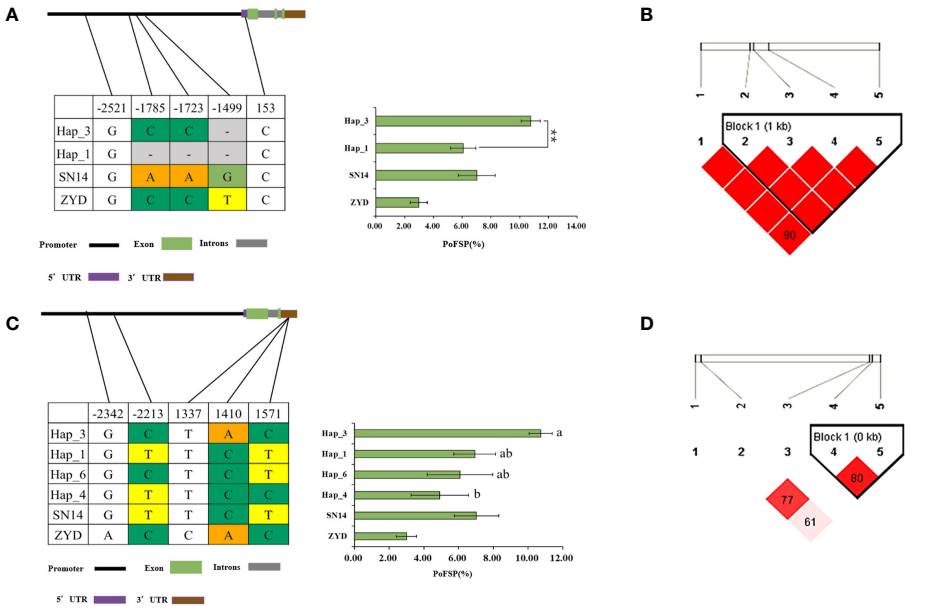


FIGURE 10
Haplotype analysis results of *Glyma.13G126400* and *Glyma.13G126700*.**(A)**: Correlation analysis of the major haplotype phenotype in *Glyma.13G126400*.**(B)**: LD analysis of SNPs among major haplotypes of*Glyma.13G126400*. **(C)**: Correlation analysis of the major haplotype phenotype in *Glyma.13G126700*.**(D)**: LD analysis of SNPs among major haplotypes of*Glyma.13G126700*. [a,c]: "**" indicates a significant difference at the 0.01 level.

preserved during the process of soybean domestication. These results are of great significance for the study of four-seed pods inheritance.

In this study, five candidate genes (*Glyma.13G126000, Glyma.13G126100, Glyma.13G126300, Glyma.13G126400, and Glyma.13G126700*) were analyzed and predicted in the fine-mapping interval. All these genes mutated between cultivated and wild parents. According to the Williams 82 reference genome gene annotation information published by the SoyBase and Phytozome websites, *Glyma.13G126000*, *Glyma.13G126100* and *Glyma.13G126300* had no annotated gene functions. This study preliminarily predicted that *Glyma.13G126000* and *Glyma.13G126100* might be negative regulators of PoFSP in soybean through the analysis of gene expression in plant lines having antagonistic extreme phenotype, but *Glyma.13G126300* might be a positive regulator of PoFSP in soybean. Additional research is needed to verify its regulatory mechanism on soybean four-seed pods. *Glyma.13G126400* belongs to the L14b domain of ribosomal proteins. Ribosomal proteins are an important part of the ribosome and have an important impact on the reproductive process of cells (Jin et al., 2018). Wu et al. (2012) found that the mitochondrial ribosomal protein GCD1 was required for the maturation of Arabidopsis female gametes (Wu et al., 2012). In 2014, Agustin et al. found that the Arabidopsis ribosomal protein L27a promoted female gametophyte development in a dose-dependent manner (Zsögön et al., 2014). Zhang et al. found in 2015 that the Arabidopsis mitochondrial ribosomal protein gene *HEART STOPPER* was required for seed development (Zhang et al., 2015). Yan et al. found in 2016 that the Arabidopsis ribosomal protein L18aB was necessary for male gametophyte and embryonic development, and Xie et al. found in 2018 that L18aB was involved in the formation of embryonic stalks in early embryonic development, further confirming that L18aB was required for reproductive development (Yan et al., 2016; Xie et al., 2018). In 2017, Lu et al. found that the Arabidopsis mitochondrial ribosomal protein S9 M was required for central cell maturation and endosperm development, and in 2020, they found that S9 M was also involved in male gametogenesis and seed development (Lu et al., 2017; Lu et al., 2020). In 2020, Xiong et al. found that the Arabidopsis ribosomal protein L27a participated in the formation of gametophytes by interacting with its molecular chaperone KETCH1 (Xiong et al., 2020). Luo et al. found in 2020 that the Arabidopsis cytoplasmic ribosomal protein L14B was required for Arabidopsis fertilization (Luo et al., 2020). Therefore, it was speculated that the *Glyma.13G126400* gene might perform important functions during the process of ovule differentiation, thereby affecting the differentiation of seed pods in soybean. The *Glyma.13G126700* gene expresses the homology domain of Gnk2. Gnk2 is a small-molecule antibacterial peptide (Sawano et al., 2007). Its role in the formation of four-seed pods and ovule differentiation needs

TABLE 4 Statistics analysis of PoFSP in the resource population.

Phenotype	Min	Mix	Mean	Standard Deviation	CV	Skewness	Kurtosis
PoFSP	0	52.78%	9.50%	12.11	1.27	1.40	1.20

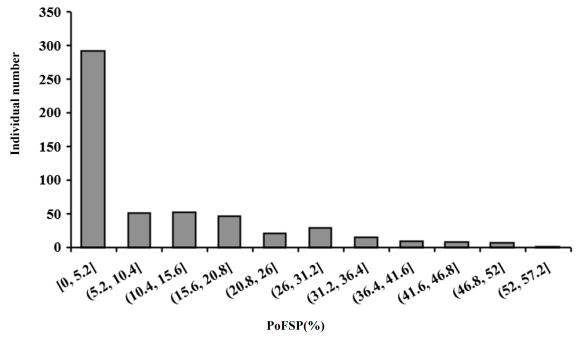


FIGURE 11
Frequency histogram of the phenotypic distribution of PoFSP in the resource population.

further research and verification. qRT-PCR results showed that both *Glyma.13G126400* and *Glyma.13G126700* might be positive regulators of PoFSP in soybean.

Haplotype analysis is an effective method to study the genetic diversity of genes using soybean germplasm resources. In this study, the resource population contained 527 varieties were used for haplotype analysis of the above five candidate genes for PoFSP. In 2021, the number of one, two, three and four-seed pods at the maturity stage of the resource population were investigated in the field, and PoFSP was calculated. In 2021, the maximum PoFSP in the resource population was 52.78%, the minimum value was 0, and the average value was 9.5%. Skewness and kurtosis were 1.4 and 1.2, respectively, and CV was 1.27 (Table 4; Figure 11). In more than half of the resource population contained 527 varieties, PoFSP was between 0 and 5%. Haplotype analysis of candidate genes for PoFSP in the resource population was helpful for the discovery of rare variation materials in PoFSP. The five candidate genes all had two or more major haplotypes in the resource population. Through significance analysis of phenotypic differences between major haplotypes, it was found that the five candidate genes all had haplotype differences in the resource population. In this study, the number of cultivars exceeded 5% of the total resource population to classify the major haplotypes. Although haplotypes with extreme phenotypes might be missed, major haplotypes could be mined more precisely for allelic diversity. According to the present results, the effects of the above five genes on soybean PoFSP were possible, but the specific effects and mechanisms of each gene on PoFSP need further study.

Conclusion

In this study, PoFSP was comapped to a QTL(*qPoFSP13-1*) on chromosome 13 in CSSLs by ICIM and BSA methods. F_2 and RHL fine mapping populations were constructed for this QTL. The QTL was fine mapped between 23.91 Mb and 24.01 Mb on chromosome 13, and the interval size was 100 Kb. A total of 14 candidate genes were included in this QTL. Through sequence analysis, expression analysis and haplotype analysis, 5 candidate genes and their effects on PoFSP were analyzed and predicted in soybean.

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 authors.

Author contributions

ChL, HJ, and QC conceived the study and designed and managed the experiments. MY,XW and YZ, provided plant

lines. FC, LH, CK, TZ and CS performed trials and collected data. FC, CaL and NW completed statistical analyses of phenotypic data. FC, RW and JX contributed to writing the paper. All authors contributed to the article and approved the submitted version.

Funding

This study was financially supported by the National Key Research and Development Program of the "14th Five Year Plan" (Accurate identification of yield traits in soybean) (2021YFD120160204), National Nature Science Foundation of China (U20A2027 and 31771882), Funding Scheme for Introducing High-level Scientific and Technological Innovation Talents by Jilin Research Institutes in 2019.China Agriculture Research System of MOF and MARA (CARS-04-PS07 and CARS-04-PS11), National Key R&D Program of China (2021YFF1001202, 2021YFD1201104-02-02), Natural Science Foundation of Heilongjiang (TD2022C003, JJ2022YX0475), Hundred-thousand and million project of Heilongjiang province for engineering and technology science' soybean breeding technology innovation and new cultivar breeding (2019ZX16B01-1).

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.1104022/full#supplementary-material

References

Alyr, M. H., Pallu, J., Sambou, A., Nguepjop, J. R., Seye, M., Tossim, H.-A., et al. (2020). Fine-mapping of a wild genomic region involved in pod and seed size reduction on chromosome A07 in peanut (Arachis hypogaea l.). *Genes* 11 (12), 1402. doi: 10.3390/genes11121402

Barik, S., Pandit, E., Mohanty, S., and Pradhan, S. (2018). QTL mapping for traits at reproductive stage drought stress in rice using single marker analysis. *ORYZA-An Int. J. Rice* 55 (1), 134–140. doi: 10.5958/2249-5266.2018.00016.4

Botero-Ramírez, A., Laperche, A., Guichard, S., Jubault, M., Gravot, A., Strelkov, S. E., et al. (2020). Clubroot symptoms and resting spore production in a doubled haploid population of oilseed rape (Brassica napus) are controlled by four main QTLs. *Front. Plant Sci.* 11, 604527. doi: 10.3389/fpls.2020.604527

Broich, S. L., and Palmer, R. G. (1980). A cluster analysis of wild and domesticated soybean phenotypes. *Euphytica* 29 (1), 23–32. doi: 10.1007/BF00037246

Dinkins, R., Keim, K., Farno, L., and Edwards, L. (2002). Expression of the narrow leaflet gene for yield and agronomic traits in soybean. *J. Heredity* 93 (5), 346–351. doi: 10.1093/jhered/93.5.346

Domingo, W. E. (1945). Leaflet shape in the soybean. *J. Agric. Res.* 70, 251.

Doyle, J. J. (1990). Isolation of plant DNA from fresh tissue. *Focus* 12, 13–15.

Fang, C., Li, W., Li, G., Wang, Z., Zhou, Z., Ma, Y., et al. (2013). Cloning of ln gene through combined approach of map-based cloning and association study in soybean. *J. Genet. Genomics* 40, 93–96. doi: 10.1016/j.jgg.2013.01.002

Gao, J., Liu, C., Jiang, H., Hu, G., and Chen, Q. (2012). QTL analysis of pod number per plant in soybean under multiple locations. *Chin. J. Oil Crop Sci.* 34 (1), 1.

Guo, W. (2004). Origins of soybean cultivation in China. *J. Nanjing Agric. Univ. (Social Sci. Edition)* 4 (1), 60–69.

Hasan, N., Choudhary, S., Naaz, N., Sharma, N., and Laskar, R. A. (2021). Recent advancements in molecular marker-assisted selection and applications in plant breeding programmes. *J. Genet. Eng. Biotechnol.* 19 (1), 1–26. doi: 10.1186/s43141-021-00231-1

He, R., Shi, Y., Zhang, J., Liang, Y., Zhang, B., Li, J., et al. (2014). QTL mapping for plant height using chromosome segment substitution lines in upland cotton. *Acta Agronomica Sin.* 40 (3), 457–465. doi: 10.3724/SP.J.1006.2014.00457

He, J., Yao, L., Pecoraro, L., Liu, C., Wang, J., Huang, L., et al. (2022). Cold stress regulates accumulation of flavonoids and terpenoids in plants by phytohormone, transcription process, functional enzyme, and epigenetics. *Crit. Rev. Biotechnol.* 18:1–18. doi: 10.1080/07388551.2022.2053056

Jiang, W., Lee, J., Jin, Y.-M., Qiao, Y., Piao, R., Jang, S. M., et al. (2011). Identification of QTLs for seed germination capability after various storage periods using two RIL populations in rice. *Molecules Cells* 31 (4), 385–392. doi: 10.1007/s10059-011-0049-z

Jiang, H., Li, C., Li, R., Li, Y., Yin, Y., Ma, Z., et al. (2020). Construction of wild soybean backcross introgression lines. *Chin. J. Oil Crop Sci.* 42 (01), 8–16. doi: 10.19802/j.issn.1007-9084.2020014

Jin, X., Cao, F., Xu, L., Yin, D., Wang, H., and Piao, L. (2017). A brief discussion on use of wild soybean to create breeding resources and varieties. *J. Northeast Agric. Sci.* 42 (01), 12–15. doi: 10.16423/j.cnki.1003-8701.2017.01.004

Jin, C., Hou, M., and Pan, Y. (2018). Research progress of ribosomal protein function in arabidopsis thaliana. *Plant Physiol. J. lant Physiol. J.* 54 (02), 203–212. doi: 10.13592/j.cnki.ppj.2017.0319

Johnson, H. W., and Bernard, R. L. (1962). Soybean genetics and breeding. *Adv. Agron.* 14, 149–221. doi: 10.1016/S0065-2113(08)60438-1

Li, C. (2014). Research progress on population mapping and localization method of soybean QTL. *Soybean Sci. Technol.* 03), 20–22.

Li, C., Jiang, H., Zhang, W., Qiu, P., Liu, C., Li, W., et al. (2008). QTL analysis of seed and pod traits in soybean. *Mol. Plant Breed.* 6 (6), 1091–1100.

Li, Y., Liu, C., Wang, N., Zhang, Z., Hou, L., Xin, D., et al. (2021). Fine mapping of a QTL locus (QNFSP07-1) and analysis of candidate genes for four-seeded pods in soybean. *Mol. Breed.* 41 (11), 1–16. doi: 10.1007/s11032-021-01265-6

Liu, M., and Fan, Q. (2021). Study on the current situation and problems of soybean consumption, production and import in China. *Grain Sci. And Technol. And Economy* 46 (06), 28–35. doi: 10.16465/j.gste.cn431252ts.20210606

Li, Y., Wang, X., Huang, S., Chen, L., Ni, Z., Wang, Y., et al. (2016). QTL mapping and epistatic interaction analysis of canopy parameters in soybean. *Mol. Plant Breed.* 14 (12), 3414–3429. doi: 10.13271/j.mpb.014.003414

Luo, A., Zhan, H., Zhang, X., Du, H., Zhang, Y., and Peng, X. (2020). Cytoplasmic ribosomal protein L14B is essential for fertilization in arabidopsis. *Plant Sci.* 292, 110394. doi: 10.1016/j.plantsci.2019.110394

Lu, C., Xie, Z., Yu, F., Tian, L., Hao, X., Wang, X., et al. (2020). Mitochondrial ribosomal protein S9M is involved in male gametogenesis and seed development in arabidopsis. *Plant Biol.* 22 (4), 655–667. doi: 10.1111/plb.13108

Lu, C., Yu, F., Tian, L., Huang, X., Tan, H., Xie, Z., et al. (2017). RPS9M, a mitochondrial ribosomal protein, is essential for central cell maturation and endosperm development in arabidopsis. *Front. Plant Sci.* 8, 2171. doi: 10.3389/fpls.2017.02171

Michelmore, R. W., Paran, I., and Kesseli, R. (1991). Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. *Proc. Natl. Acad. Sci.* 88 (21), 9828–9832. doi: 10.1073/pnas.88.21.9828

Ning, H., Yuan, J., Dong, Q., Li, W., Xue, H., Wang, Y., et al. (2018). Identification of QTLs related to the vertical distribution and seed-set of pod number in soybean [Glycine max (L.) merri]. *PLoS One* 13 (4), e0195830. doi: 10.1371/journal.pone.0195830

Paterson, A. H., Lander, E. S., Hewitt, J. D., Peterson, S., Lincoln, S. E., and Tanksley, S. D. (1988). Resolution of quantitative traits into mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. *Nature* 335 (6192), 721–726. doi: 10.1038/335721a0

Peng, Y., Zhu, J., Yang, G., and Yuan, J. (1994). Relation of soybean leaf shape distribution to 4-seeded pods. *Acta Agronomica Sin.* 20 (4), 501–503.

Poretska, O., Yang, S., Pitorre, D., Poppenberger, B., and Sieberer, T. (2020). AMP1 and CYP78A5/7 act through a common pathway to govern cell fate maintenance in arabidopsis thaliana. *PLoS Genet.* 16 (9), e1009043. doi: 10.1371/journal.pgen.1009043

Qiao, W., Qi, L., Cheng, Z., Su, L., Li, J., Sun, Y., et al. (2016). Development and characterization of chromosome segment substitution lines derived from oryza rufipogon in the genetic background of o. sativa spp. indica cultivar 9311. *BMC Genomics* 17 (1), 1–12. doi: 10.1186/s12864-016-2987-5

Sawano, Y., Miyakawa, T., Yamazaki, H., Tanokura, M., and Hatano, K.-i. (2007). Purification, characterization, and molecular gene cloning of an antifungal protein from ginkgo biloba seeds. 388 (3), 273-280. doi: 10.2210/pdb2e79/pdb

Si, W., and Han, T. (2021). China's soybean yield increase potential and realization path during the "14th Five–Year Plan" Period. *Agric. Economic Issues* 07, 17–24. doi: 10.13246/j.cnki.iae.2021.07.003

Song, H., Cui, X., Ma, H., Zhu, Z., and Luan, F. (2012). Construction of DNA fingerprint database based on SSR Marker for Varieties(Lines) of cucumis melo l. *Scientia Agricultura Sin.* 45 (13), 2676–2689.

Song, X., Wang, K., Guo, W., Zhang, J., and Zhang, T. (2005). A comparison of genetic maps constructed from haploid and BC1 mapping populations from the same crossing between gossypium hirsutum l. and gossypium barbadense l. *Genome* 48 (3), 378–390. doi: 10.1139/g04-126

Sun, Y., Shi, X., Jiang, H., Sun, D., Xin, D., Liu, C., et al. (2012). Epistatic effects and qE interaction effects of QTLs for 100 – seed weight in soybean. *Chin. J. Oil Crop Sci.* 34 (06), 598–603.

Takahashi, N. (1934). Linkage relation between the genes for the forms of leaves and the number of seeds per pod of soybeans. *Jpn. J. Genet.* 9, 208–225. doi: 10.1266/jjg.9.208

Wakchaure, R., Ganguly, S., Praveen, P., Kumar, A., Sharma, S., and Mahajan, T. (2015). Marker assisted selection (MAS) in animal breeding: a review. *J. Drug Metab. Toxicol.* 6 (5), e127. doi: 10.4172/2157-7609.1000e127

Wang, N., Jiang, H., Zhang, Z., Li, Y., Li, C., Song, Y., et al. (2022). Gene mining of 100-grain weight and the number of four-seed pods in soybean (Glycine max). *Plant Breed.* 141 (2), 143–158. doi: 10.1111/pbr.12993

Wang, X., Li, Y., Zhang, H., Sun, G., Zhang, W., and Qiu, L. (2015). Evolution and association analysis of GmCYP78A10 gene with seed size/weight and pod number in soybean. *Mol. Biol. Rep.* 42 (2), 489–496. doi: 10.1007/s11033-014-3792-3

Wang, X., Zhang, X., Zhou, R., Sha, A., Wu, X., Cai, S., et al. (2007). QTL analysis of seed and pod traits in soybean RIL population. *Acta AGRONOMICA Sin.* 33 (3), 441–448.

Weiss, M. G. (1970). Genetic linkage in soybeans: linkage group I 1. *Crop Sci.* 10 (1), 69–72. doi: 10.2135/cropsci1970.0011183X001000010027x

Wu, J.-J., Peng, X.-B., Li, W.-W., He, R., Xin, H.-P., and Sun, M.-X. (2012). Mitochondrial GCD1 dysfunction reveals reciprocal cell-to-cell signaling during the maturation of arabidopsis female gametes. *Dev. Cell* 23 (5), 1043–1058. doi: 10.1016/j.devcel.2012.09.011

Xie, J., Wang, Q., Zhang, Z., Xiong, X., Yang, M., Qi, Z., et al. (2021). QTL-seq identified QTLs and candidate genes for two-seed pod length and width in soybean (Glycine max). *Plant Breed.* 140 (3), 453–463. doi: 10.1111/pbr.12920

Xie, F., Yan, H., Sun, Y., Wang, Y., Chen, H., Mao, W., et al. (2018). RPL18aB helps maintain suspensor identity during early embryogenesis. *J. Integr. Plant Biol.* 60 (4), 266–269. doi: 10.1111/jipb.12616

Xiong, F., Duan, C.-Y., Liu, H.-H., Wu, J.-H., Zhang, Z.-H., Li, S., et al. (2020). Arabidopsis KETCH1 is critical for the nuclear accumulation of ribosomal proteins and gametogenesis. *Plant Cell* 32 (4), 1270–1284. doi: 10.1105/tpc.19.00791

Xi, Z., Zhu, F., Tai, G., and Li, Z. (2005). Principle and methodology of QTL analysis in crop. *Chin. Agric. Sci. Bull.* 01), 88–92+99.

Xu, Z., Wang, F., Zhou, Z., Meng, Q., Chen, Y., Han, X., et al. (2022). Identification and fine-mapping of a Novel QTL, qMrdd2, that confers resistance to maize rough dwarf disease. *Plant Dis.* 106 (1), 65–72. doi: 10.1094/PDIS-03-20-0495-RE

Yan, H., Chen, D., Wang, Y., Sun, Y., Zhao, J., Sun, M., et al. (2016). Ribosomal protein L18aB is required for both male gametophyte function and embryo development in arabidopsis. *Sci. Rep.* 6 (1), 1–12. doi: 10.1038/srep31195

Yang, Z., Huang, D., Tang, W., Zheng, Y., Liang, K., Cutler, A. J., et al. (2013a). Mapping of quantitative trait loci underlying cold tolerance in rice seedlings *via* high-throughput sequencing of pooled extremes. *PLoS One* 8 (7), e68433. doi: 10.1371/journal.pone.0068433

Yang, Z., Sun, Y., Qi, Z., Xin, D., Jiang, H., He, L., et al. (2013b). Analysis of additive effect,epistatic and QE interaction effect for QTL of pod number traits in soybean. *J. Agric. Univ.* 3), 1–13.

Yao, D., Liu, Z., Zhang, J., Liu, S., Qu, J., Guan, S., et al. (2015). Analysis of quantitative trait loci for main plant traits in soybean. *Genet. Mol. Res.* 14 (2), 6101–6109. doi: 10.4238/2015.June.8.8

Yuan, R., Huang, Z., Luo, L., Zhao, N., Chen, Y., Liang, Y., et al. (2020). Identification of grain-shattering QTL and preliminary mapping of a related major QTL based on chromosome segment substitution lines (CSSLs) of guangxi common wild rice (Oryza rufipogon griff.). *J. South. Agric.* 51 (5), 1004–1012. doi: 10.3969/j.issn.2095-1191.2020.05.003

Zhang, H., Luo, M., Day, R. C., Talbot, M. J., Ivanova, A., Ashton, A. R., et al. (2015). Developmentally regulated HEART STOPPER, a mitochondrially targeted L18 ribosomal protein gene, is required for cell division, differentiation, and seed development in arabidopsis. *J. Exp. Bot.* 66 (19), 5867–5880. doi: 10.1093/jxb/erv296

Zhang, X., Wang, W., Guo, N., Zhang, Y., Bu, Y., Zhao, J., et al. (2018). Combining QTL-seq and linkage mapping to fine map a wild soybean allele characteristic of greater plant height. *BMC Genomics* 19 (1), 1–12. doi: 10.1186/s12864-018-4582-4

Zhao, L., Zhou, H., Lu, L., Liu, L., Li, X., Lin, Y., et al. (2009). Identification of quantitative trait loci controlling rice mature seed culturability using chromosomal segment substitution lines. *Plant Cell Rep.* 28 (2), 247–256. doi: 10.1007/s00299-008-0641-7

Zheng, H., Hou, L., Xie, J., Cao, F., Wei, R., Yang, M., et al. (2022). Construction of chromosome segment substitution lines and inheritance of seed-pod characteristics in wild soybean. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.869455

Zhou, L., Wang, S.-B., Jian, J., Geng, Q.-C., Wen, J., Song, Q., et al. (2015). Identification of domestication-related loci associated with flowering time and seed size in soybean with the RAD-seq genotyping method. *Sci. Rep.* 5 (1), 1–8. doi: 10.1038/srep09350

Zhuo, R., Chen, H., Wang, X., Zhang, X., Shan, Z., Wu, X., et al. (2009). QTL analysis of yield, yield components, and lodging in soybean. *Acta AGRONOMICA Sin.* 35 (5), 821–830. doi: 10.3724/SP.J.1006.2009.00821

Zou, C., Wang, P., and Xu, Y. (2016). Bulked sample analysis in genetics, genomics and crop improvement. *Plant Biotechnol. J.* 14 (10), 1941–1955. doi: 10.1111/pbi.12559

Zsögön, A., Szakonyi, D., Shi, X., and Byrne, M. E. (2014). Ribosomal protein RPL27a promotes female gametophyte development in a dose-dependent manner. *Plant Physiol.* 165 (3), 1133–1143. doi: 10.1104/pp.114.241778

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 15 February 2023
DOI 10.3389/fpls.2023.1091432

OPEN ACCESS

EDITED BY
Kyuya Harada,
Osaka University, Japan

REVIEWED BY
Xin Li,
Nanjing Agricultural University, China
Mahendar Thudi,
Dr. Rajendra Prasad Central Agricultural University, India

*CORRESPONDENCE
Gyan P. Mishra
gyan.gene@gmail.com
Harsh Kumar Dikshit
harshgeneticsiari@gmail.com
Shiv Kumar
SK.Agrawal@cgiar.org

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Functional and Applied Plant Genomics, a section of the journal Frontiers in Plant Science

RECEIVED 07 November 2022
ACCEPTED 31 January 2023
PUBLISHED 15 February 2023

CITATION
Dutta H, K. M. S, Aski MS, Mishra GP, Sinha SK, Vijay D, C. T. MP, Das S, Pawar PA-M, Mishra DC, Singh AK, Kumar A, Tripathi K, Kumar RR, Gupta S, Kumar S and Dikshit HK (2023) Morpho-biochemical characterization of a RIL population for seed parameters and identification of candidate genes regulating seed size trait in lentil (*Lens culinaris* Medik.).
Front. Plant Sci. 14:1091432.
doi: 10.3389/fpls.2023.1091432

Morpho-biochemical characterization of a RIL population for seed parameters and identification of candidate genes regulating seed size trait in lentil (*Lens culinaris* Medik.)

Haragopal Dutta[1†], Shivaprasad K. M.[1†], Muraleedhar S. Aski[1†], Gyan P. Mishra[1*], Subodh Kumar Sinha[2], Dunna Vijay[3], Manjunath Prasad C. T.[3], Shouvik Das[4], Prashant Anupama-Mohan Pawar[4], Dwijesh C. Mishra[5], Amit Kumar Singh[6], Atul Kumar[3], Kuldeep Tripathi[7], Ranjeet Ranjan Kumar[8], Sanjeev Gupta[9], Shiv Kumar[10*] and Harsh Kumar Dikshit[1*]

[1]Division of Genetics, Indian Agricultural Research Institute, New Delhi, India, [2]Indian Council of Agricultural Research (ICAR)-National Institute for Plant Biotechnology, New Delhi, India, [3]Division of Seed Science and Technology, Indian Agricultural Research Institute, New Delhi, India, [4]Laboratory of Plant Cell Wall Biology, Regional Centre for Biotechnology, Faridabad, India, [5]Agricultural Bioinformatics, Indian Agricultural Statistics Research Institute, New Delhi, India, [6]Division of Genomic Resources, National Bureau of Plant Genetic Resources, New Delhi, India, [7]Germplasm Evaluation Division, National Bureau of Plant Genetic Resources, New Delhi, India, [8]Division of Biochemistry, Indian Agricultural Research Institute, New Delhi, India, [9]Krishi Bhawan, Indian Council of Agricultural Research, New Delhi, India, [10]South Asia and China Program, International Center for Agricultural Research in the Dry Areas, National Agriculture Science Complex (NASC) Complex, New Delhi, India

The seed size and shape in lentil (*Lens culinaris* Medik.) are important quality traits as these influences the milled grain yield, cooking time, and market class of the grains. Linkage analysis was done for seed size in a RIL ($F_{5:6}$) population derived by crossing L830 (20.9 g/1000 seeds) with L4602 (42.13 g/1000 seeds) which consisted of 188 lines (15.0 to 40.5 g/1000 seeds). Parental polymorphism survey using 394 SSRs identified 31 polymorphic primers, which were used for the bulked segregant analysis (BSA). Marker PBALC449 differentiated the parents and small seed size bulk only, whereas large seeded bulk or the individual plants constituting the large-seeded bulk could not be differentiated. Single plant analysis identified only six recombinant and 13 heterozygotes, of 93 small-seeded RILs (<24.0 g/1000 seed). This clearly showed that the small seed size trait is very strongly regulated by the locus near PBLAC449; whereas, large seed size trait seems governed by more than one locus. The PCR amplified products from the PBLAC449 marker (149bp from L4602 and 131bp from L830) were cloned, sequenced and BLAST searched using the lentil reference genome and was found amplified from chromosome 03. Afterward, the nearby region on chromosome 3 was searched, and a few candidate genes like ubiquitin carboxyl-terminal hydrolase, E3 ubiquitin ligase, TIFY-like protein, and hexosyltransferase having a role in seed size determination

were identified. Validation study in another RIL mapping population which is differing for seed size, showed a number of SNPs and InDels among these genes when studied using whole genome resequencing (WGRS) approach. Biochemical parameters like cellulose, lignin, and xylose content showed no significant differences between parents and the extreme RILs, at maturity. Various seed morphological traits like area, length, width, compactness, volume, perimeter, etc., when measured using VideometerLab 4.0 showed significant differences for the parents and RILs. The results have ultimately helped in better understanding the region regulating the seed size trait in genomically less explored crops like lentils.

KEYWORDS

BSA, cell membrane, cellulose, lignin, masur, videometer, xylose

Introduction

Lentil (*Lens culinaris* ssp. *culinaris* Medik.) is a diploid (2n=14), self-pollinated, cool season legume crop having a genome size of nearly 4.2 Gb (Arumuganathan and Earle, 1991; Dikshit et al., 2022a; Dikshit et al., 2022b). This is not only rich in proteins but also in micronutrients (Fe and Zn) and β-carotene (Mishra et al., 2020; Priti et al., 2021; Priti et al., 2022). Lentil is being grown throughout the world in temperate to sub-tropical regions including regions of the Middle East, north-eastern Africa, Southern Europe, South and North America, Australia, and the Indian sub-continent (Mishra et al., 2022a). Globally, Canada is the largest producer and exporter of lentils. Lentil is an important crop for India having acreage of 1.35 m ha and production of 1.18 m tons. The world production of lentils is 6.54 m tons from an area of nearly 5.01 m ha. Lentil productivity in India (871.5 kg/ha) is well below world productivity (1304.9 kg/ha) (FAOSTAT, 2020).

Seed quality of lentil is an important objective for both industry and the consumer. Among various parameters, seed size is the key parameter defining the overall lentil quality (Singh et al., 2022). During domestication of lentils, several traits like pod dehiscence, dormancy, and seed size got modified which ultimately allowed easy collection of seeds by the farmers for next year sowing (Sonnante et al., 2009). Most of the domestication traits like pod dehiscence, dormancy, and growth habit are single gene governed traits while seed size is a quantitative trait. Depending upon the seed size lentil is classified into microsperma type (2 to 6 mm diameter, red and yellow cotyledons, and pigmented flowers) and macrosperma type (6 to 9 mm diameter, yellow cotyledon, and non-pigmented flowers) (Barulina, 1930; Sandhu and Singh, 2007). Generally, microsperma types are more common in southeast Asia, while macrosperma types in western Asia and Europe (Barulina, 1930). Previous genetic studies revealed large variations for seed weight and seed diameter in lentils (Tullu et al., 2001; Dutta et al., 2022; Tripathi et al., 2022).

Seed size and shape are known to influence both cooking time and dehulling efficiency and are considered important market-associated trait (Erskine et al., 1991; Wang, 2008). A strong positive correlation (r=0.96) was recorded between seed size and cooking time (Hamdi et al., 1991). Ford et al. (2007) noted reduced damage during handling in the rounder seed-shaped lentil cultivars over thin, sharp-edged types. Thus, the development of genotypes with improved seed parameters including seed weight is an important breeding objective of lentil breeders across the globe (Mishra et al., 2022b). Generally, seed parameters are measured using crude phenotypic evaluation methods like measurement of 100-grain weight or seed diameter measurement using Vernier caliper or graded sieve (Hossain et al., 2010; Xu et al., 2011). In lentils, seed diameter was also measured using computer-assisted 2-dimensional imaging (Shahin and Symons, 2001) and seed plumpness was determined using 3-dimensional imaging using a camera (Shahin et al., 2006). However, these are laborious methods, especially when a large number of genotypes are involved in screening. Recently, Dutta et al. (2022) used a very easy method involving VideometerLab 4.0 instrument for the measurement of various seed parameters in lentils.

Linked molecular markers with the trait of interest will help in efficient breeding for that trait (Mishra et al., 2003). Seed weight is known to be governed by several genes and thus identification of linked markers with the seed weight QTLs will help in the better selection for this trait. This will also help in speeding up of new variety development having desired seed parameters (Tripathi et al., 2022). Also, for the implementation of molecular breeding approach for seed size trait, there is a need for the development of an experimental population involving contrasting parents, so that the linkage can be established between marker and the trait. Evaluation of a RIL population with SSRs markers using BSA approach can help in the identification of linked markers with the seed size trait in lentil (Mishra et al., 2001; Mishra et al., 2003). This study hypothesizes that the genomic region controlling the seed size can be identified using molecular markers in a mapping population differing for the seed size trait. Against this backdrop, the objective of this study was to perform the morpho-biochemical characterization of a RIL population for seed parameters and identification of candidate genes regulating seed size trait in lentil.

Materials and methods

Plant materials

Two lentil genotypes differing significantly in seed size, L830 (small-seeded; mean 1000 seed weight = 20.9 g) and L4602 (large-seeded; mean 1000 seed weight = 42.13g) were used as the parent for the development of a RIL population (Figure 1). Cross was attempted between the L4602 and L830 and the F_1 was confirmed for its hybridity using polymorphic SSR markers. The parents and the RIL population ($F_{5:6}$) having 188 lines were grown during rabi-2021 at the fields of Indian Agricultural Research Institute, New Delhi, India (Latitude: 28.6412°N, Longitude: 77.1627°E, and Altitude: 228.61 m AMSL) with the spacing of 30×5 cm (row to row × plant to plant) in a 5.0 m row length using standard cultivation practices. Each row was harvested at maturity and 1000 seed weight was measured for parents and the RILs (Figures 2, S1).

DNA extraction and constitution of bulks for bulked segregant analysis

Nearly 15-20 seeds each from 188 RILs along with the parents (L830 and L4602) were kept on the germination paper and was wrapped in a butter paper. This was then kept in a germination chamber for 8-10 days at 20-25°C. The tender seedlings were used for DNA isolation using CTAB method (Murray and Thompson, 1980) and quality was checked on 0.8% Agarose gel, while quantity was measured using Nanodrop (García-Alegría et al., 2020). A total of 10 extreme genotypes each from small-seeded RILs (line No. 05, 14, 16, 64, 88, 111, 117, 155, 160, 169) and large-seeded RILs (line No. 03, 86, 87, 97, 102, 107, 108, 115, 133, 190) were used for the BSA (Figure 3). An equal quantity of DNA (20 ng/µL) was taken from each of the 10 extreme RILs and mixed to constitute the two contrasting bulks (B1 and B2). A total of 394 SSRs were used for the parental polymorphism survey (Table S1) and polymorphic primers were used for BSA (Michelmore et al., 1991) and band were separated on 3.0% Metaphor agarose gel and scored. The SSRs differentiating the bulks and the parents were used for the individual RIL analysis. The RILs were arranged in the increasing order of their seed size, PCR was performed and amplification was visualized on the gel using gel documentation system.

Cloning and sequencing of a PCR amplified product

The DNA fragment amplified by an SSR marker (PBALC449) in L4602 and L830 was used for the cloning and sequencing. The amplified bands containing the DNA were first precisely excised from the gel with a clean, sharp scalpel and then DNA was extracted using QIAquick Gel Extraction Kit (QIAGEN, Valencia, USA) by following the manufacturer's instructions (Sambrook et al., 1989). The amplified product was ligated with pJET1.2/blunt vector using CloneJET PCR Cloning Kit (Thermo Fisher Scientific™) as per the mentioned protocol (https://www.thermofisher.com/document-connect/document-connect.html). Then the recombinant vector was transformed into *E. coli* DH5α strain competent cells for cloning using the standard protocol. Afterward, plasmid was isolated using FavorPrep Plasmid Extraction Mini Kit as per the manufacturer's instructions and extracted plasmid DNA was stored at -20°C for further analysis. The cloning was confirmed by restriction digestion using *Bgl* II. The positive clones were sequenced using Sanger sequencing platform using universal primer. The raw sequence data was processed by trimming the vector sequence and aligned to the reference genome (CDC Redberry Genome Assembly v2.0; Ramsay et al., 2021) using NCBI BLAST browser.

Biochemical analysis of lentil genotypes and the extreme RILs differing for seed size

Various cell wall related biochemical analyses were performed on the 10 extreme RILs each for seed size (large and small seeded RILs), and the parents.

FIGURE 1
Seeds of the small-seeded (L830, mean 1000 seed weight=20.9g) and large-seeded (L4602, mean 1000 seed weight=42.13g) parents were used for the seed size analysis.

FIGURE 2
A representative figure showing variations for the seed parameters in a set of 50 RILs ($F_{5:6}$), derived from the cross L4602×L830 (mean 1000 seed weight range=15.0 to 40.5g).

Preparation of alcohol insoluble residue sample

Briefly, 600 mg of lentil seeds were crushed, flash frozen (in liquid N_2), and ground in Qiagen TissueLyser II (at 30 Hz for 2-3 min) to a fine powder. Then 100 mg powder was taken for incubation (at 70°C for 30 min) in 5.0 mL ethanol (80%) having 4.0 mm HEPES buffer (pH 7.5). This was then cooled on ice and centrifuged (1000 rpm for 15.0 min), supernatant was discarded, residue was washed (5.0 mL 70% ethanol) and then suspended in chloroform: methanol (1:1) solution (5.0 mL) for 3.0 min at room temperature and centrifuged (14000 rpm for 15.0 min). The residue was again washed with acetone (5.0 mL), pellet was dried in a desiccator and used as an AIR sample for further analysis (Pawar et al., 2017).

Estimation of cellulose by Updegraff method

To the AIR sample (2.0 mg), Updegraff reagent (acetic acid: nitric acid: water = 8:1:2 v/v) was added and incubated (at 100°C for 30 min). The mixture was then centrifuged (10,000 rpm; 15 min), and pellet was washed four times with acetone and dried overnight. The dried residue was hydrolyzed in 72% H_2SO_4, glucose was analyzed by anthrone assay (Updegraff, 1969) and a standard curve was used to estimate the cellulose content.

Estimation of xylose and O-acetyl content

AIR sample (2.0 g) was incubated for neutralization with HCl and NaOH for xylose and acetyl content estimation, respectively. The xylose and O-acetyl content were analyzed using Megazyme K-ACET and K-XYLOSE kits, respectively (Rastogi et al., 2022).

Acetyl bromide soluble lignin content

The 25% acetyl bromide solution was diluted using acetic acid and incubated at 50°C for 2.0h. The solubilized powder was mixed with NaOH and hydroxylamine hydrochloride and then absorbance was recorded at 280 nm and lignin content was measured (Foster et al., 2010).

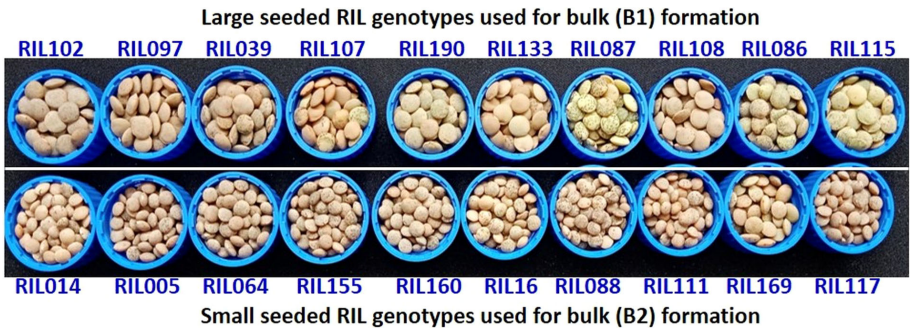


FIGURE 3
The 10 extreme RIL genotypes for seed size which was used for the formation of two extreme bulks for the BSA. Where, upper panel represents 10 RILs with maximum seed size (in descending order of their seed weight) and lower panel represents 10 RILs with minimum seed size (in ascending order of their seed weight).

Lignin and cellulose estimation through fourier transform-infrared spectroscopy

Lignin and cellulose contents were estimated using FTIR spectroscopy in the lentil seed powder (Pawar et al., 2017). A Tensor FTIR spectrometer (Bruker Optics) equipped with a single-reflectance horizontal ATR cell (ZnSe Optical Crystal, Bruker Optics) was used for the analysis. The spectrum range selected was from 600 cm^{-1} to 4000 cm^{-1} having a resolution of 4 cm^{-1}. KBr powder was used for the preparation of standard and each sample was measured twice (by removing and adding different aliquots of powder for heterogeneity evaluation) and each spectrum was the average of 16 scans (Labbe et al., 2005; Canteri et al., 2019).

Estimation of seed morphological parameters using VideometerLab 4.0 instrument

Detailed seed phenotyping was done for eight RILs (four large and small-seeded each) and the parents using VideometerLab 4.0 instrument (Videometer A/S, Denmark) which captured the images of 30 seeds placed in a customized 3D printed plate. Videometer acquires morphological and spectral information using 19 high power LED sources (375, 405, 435, 450, 470, 505, 525, 570, 590, 630, 645, 660, 700, 780, 850, 870, 890, 940, 970 nm). The data were quantified using custom-designed software (VideometerLab software ver. 2.13.83) as seed area, length, width, etc. (Shrestha et al., 2015).

Statistical analysis

ANOVA was performed to determine the genotypic variance among parents and the 188 RIL genotypes (for various seed parameters like seed weight, area, length, width, width/length, compactness, width/area, volume, and perimeter) and also among the parents and the 10 extreme RIL genotypes (for biochemical parameters like lignin, cellulose, and xylose contents) using DSAASTAT ver.1.514 software. Afterward, multiple comparison test was performed using Tukey HSD method ($p \leq 0.05$).

Results

Identification of linked marker(s) with seed size in RIL population using BSA

For parental polymorphism, 394 SSR primer pairs of different series like PBALC (Kaur et al., 2011), PLC (Jain et al., 2013), LC (Verma et al., 2014), and GLLC (Saha et al., 2010) have been used, and 31 were found polymorphic (Tables 1, S1; Figure S2). The bulked segregant analysis (BSA) was performed on the parents and the two bulks made by mixing equal quantity of DNA from the 10 extreme RIL genotypes for seed size using polymorphic markers (Figure 3). Of 31 polymorphic SSRs, only one PBLAC449, clearly differentiated the small seed size bulk and the parent, whereas large seed size bulk showed two bands. However, other polymorphic markers could not differentiate the bulk (Figure 4). The PBLAC449 primer amplified 131bp band for L830 and 149bp band for L4602 genotype. To understand this unique type of banding pattern, the individual plants constituting the bulk was amplified. As observed for the bulked samples, all the 10 plants of small seed size samples showed a band similar to the small seed parent i.e. L830 (131 bp). However, the individual plants constituting the large seed size bulk, a mix of amplification patterns with 03 recombinants (having L830 band size) and 02 heterozygotes were recorded (Figure 5).

Afterward, DNA of RILs were rearranged in the order of increasing seed weight, and then PCR (and gel electrophoresis) was done using PBALC449 primer. Of 188 RILs, 39 lines showed 149 bp amplification (as L4602 type), 43 lines showed heterozygous (both 149 and 131 bp bands), and 106 lines showed 131 bp amplification (as L830 type) (Figure S3, Table S2). Based on the seed size, the RILs were broadly grouped into two categories (i) >24.0 g/1000 seeds (large seeded; 95 Numbers) and (ii) <24.0g/1000 seeds (small seeded; 93 Numbers), expecting that the major locus must have been fixed in a 1:1 ratio in the RILs. Of 188 RILs, the first 73 RILs (15.0 to 21.4 g/1000 seed) which were arranged in the increasing order of seed weight, showed only 03 recombinants (and 07 heterozygotes), while the first 93 lines showed only six recombinants (and 13 heterozygotes). This kind of unique banding pattern has clearly suggested the presence of very tight linkage between small seed size trait and PBLAC449 marker and also indicated that there is no marker distortion in the studied population (Table 2).

Interestingly, this marker showed independent segregation for the large-seeded trait. Thus, it seems that the large seed size expression is being governed by two or more major loci. Since the banding pattern was so unique that we were unable to use any standard marker linkage analysis. To decipher such a unique type of banding pattern, we decided to find the chromosomal location of the amplified product (tightly linked with the small seed size trait only) by cloning, sequencing, and the comparative genomics approach.

Cloning, sequencing and chromosomal location of PCR amplified products from PBALC449 marker

The pJET1.2 vector was used for cloning of the PCR amplified products (149bp from L4602 and 131bp from L830) from a putatively linked marker viz., PBALC449 for small seed size trait in lentil. The cloned fragment was then sequenced which was further aligned to the reference genome (CDC Redberry Genome Assembly v2.0) using NCBI BLAST browser. The difference in the total length of the amplified product between both parents was due to the presence of 18bp deletion at two places (Table S3). The alignment details of the amplified product with the reference genome is presented in Figures S4-S5. The position of PBALC449 amplification was at Luc.2RBY.Chr3:398437705.398441563 (+strand) which is a PsbP domain protein-encoding gene (3859 bp) and is present on chromosome number 3 of lentil (Figures 6, S6). The related species

10.3389/fpls.2023.1091432

TABLE 1 Details of the SSRs found polymorphic between the parents L4602 and L830.

S. No	Primer name	Forward sequence (5′-3′)	Reverse Sequence (5′-3′)	Product size (bp)	Motif	Reference
1.	PBALC 114	CACCATAGTGACTACCACCAC	GACAGTGAGGTTGTTGAAAAG	151	(ACC)4	Kaur et al., 2011
2.	PBALC 209	GGAGTTGGTTAGAAGGAAAGA	CTAGATATCATCGATCCATCC	152	(GTC)4	Kaur et al., 2011
3.	PBALC 449	CAGCAATGGTTTTACACTCTC	GGATTTGTTTTGGTTAAGGAT	149	AAC	Kaur et al., 2011
4.	PBALC 761	GTTTGTTATCGTTGGAAGGTT	GAAGCTTAGTGAGAGCAAAAGT	156	CTT	Kaur et al., 2011
5.	PLC 34	TACTGGATGAGACGAAGATGGA	CGAAACCTGGCCTATACAAAAG	190	(T)10	Jain et al., 2013
6.	PLC 36	ACTCAAGTCAACCTCAGAAGGC	CTTAGGAGCCGGAGAAGAAGAT	500	(CTTCA)3	Jain et al., 2013
7.	PLC 37	CTCTCCAGTCCTTGCTTGATG	ACCAACAAACTTGCCAGACTTC	100	(T)13	Jain et al., 2013
8.	PLC 42	AACCAATCATGGCTTCTGCT	TTTCACCGTCTTTATGAACCA	220	(GA)8	Jain et al., 2013
9.	PLC 44	AAATGGTGCATGTGTACGGT	GGAGAACGCGATCAGTAAGG	110	(GCC)5	Jain et al., 2013
10.	PLC 45	CCTTAGTCACTGTGGTCTGATGA	ACAATGAGAGGCCAGTGCTT	390	(A)12	Jain et al., 2013
11.	PLC 51	CCATGATGAGCCTTGAATGA	TCTTCAATCTCCAGGAACACTTT	120	(GAA)10	Jain et al., 2013
12.	PLC 60	TGCTTGGACCCTAAATTTGC	AAGAAAAGGGCAACCACTGA	190	(TA)6	Jain et al., 2013
13.	PLC 66	ATTTGGAGCAAAGATGCAGG	GGATCGACCTCCAATCAAGA	340	(A)10	Jain et al., 2013
14.	PLC 69	CGCTCTACCAACAGCATAA	GAGGTCTCTTTTGTTCTTCACT	210	(CT)19	Jain et al., 2013
15.	PLC 70	CATCTCTTCGTGGCGTAAT	AGCAAACAACAGCACACATA	250	(GTT)9	Jain et al., 2013
16.	PLC 77	GGAAAGAGCCAAGAAGTTG	ACCCATCCTCATCCTTAAAT	230	(CAATGG)5	Jain et al., 2013
17.	PLC 80	GCTAACAAACAACACCATGA	GCATCTAAGTTCTTCAATCTCC	185	(GAA)10	Jain et al., 2013
18.	PLC 105	CTCCCTCAAAATGCGTTGAT	TCCATTACAAGATACTCTCCATGC	320	(TTTTA)6	Jain et al., 2013
19.	LC 272	CAAGATTCCGCACCAATACG	GTTCGGGGGTAATCCAAACT	250	(CT)4	Verma et al., 2014
20.	LC 301	GCCCTAAGTCACCAGAAAACA	CCCTTCGAACCATAATCGTG	300	(GA)4…(GA)5	Verma et al., 2014
21.	LC 305	ACTATTAGCGAAGCCCAGCA	TGAATCCAGAGCCTTTCTTTG	400	CT	Verma et al., 2014
22.	LC 307	AAGTCGACCTTATGAATGAGCA	CAGAACACTGCGAGGTATGA	471	(GA)8…(GA)9	Verma et al., 2014
23.	LC 385	GCCTTTTCAACAGCTACTTTGTT	TGCTTGAGAAATCTGACACACA	392	(CT)7….(GAA)4	Verma et al., 2014
24.	LC 389	TGTCAGCGTAAGATTGGACA	GCAAAGATTTGCTTCAACAAG	384	(CAA)3…(CT)8	Verma et al., 2014
25.	LC 396	GGTCTCTCAAGACTATTGCAAGAAA	TGGATCAAGTGGTATATTTGGACA	292	(AA)3…(GA) 10.(CTTT)2	Verma et al., 2014

(Continued)

TABLE 1 Continued

S. No	Primer name	Forward sequence (5′-3′)	Reverse Sequence (5′-3′)	Product size (bp)	Motif	Reference
26.	LC 398	TTGTGGTCACTCAAGACTATTGC	CAAGACTACTCTAGCCTTTTCAACG	392	(CTT)4…(GA)13	Verma et al., 2014
27.	LC 421	CTTTCTTTGAATATGAACGTGAGAG	GCCTTTTCAACGGCTCCT	250	(GA)8	Verma et al., 2014
28.	GLLC 541	TGGGCTCATTGAACCAAAAG	CCCCCTTTTAAGTGATTTTCC	450	–	Saha et al., 2010
29.	GLLC 562	TGTGTAGGCACATCAACAAAA	GGTGGGCATGAGAGGTGTTA	420	–	Saha et al., 2010
30.	GLLC 563	ATGGGCTCATTGAACAAAAG	CCCCCTCTAAGAGATTTTCCTC	300	–	Saha et al., 2010
31.	GLLC 614	AACCCCAGCCAGATCTTACA	AAGGGTGGTTTTGGTCCTATG	–	–	Saha et al., 2010

sequence similarity showed maximum similarity with *Medicago truncatula* and was followed by *Cicer arietinum* (Figure S7). To identify the candidate genes regulating small seed size trait near this marker, we checked RNA Seq data generated by us using the same parental combinations (Dutta et al., 2022) and the physical chromosomal details available at CDC Redberry Genome Assembly v2.0 (Ramsay et al., 2021). Using KnowPulse browser, on the left side (0.6 Mb region) of the PBALC449 amplified region, three candidate genes namely, E3 Ubiquitin ligase (log2FC -1.582), TIFY-like protein, and hexosyltransferase gene (log2FC -2.474); while on the right side (in 0.7 Mb region), a ubiquitin carboxyl-terminal hydrolase gene was found (Ramsay et al., 2021).

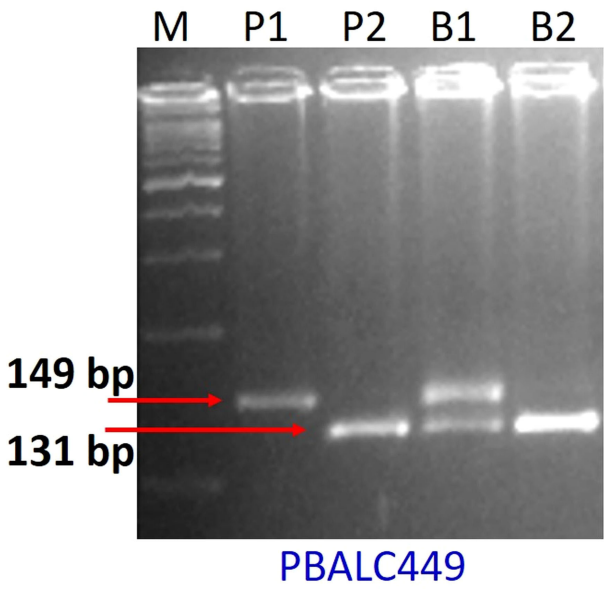


FIGURE 4
Gel picture showing BSA for seed size with PBALC449 marker. Where, P1: L4602, P2: L830, B1: large seeded bulk, B2: small-seeded bulk, M: DNA ladder.

Estimation of lignin, cellulose, xylose, and acetyl content in the parents and the 10 extreme RILs differing for the seed size

Cell wall composition is known to determine the size and shape of some seeds. To validate this, we analyzed and compared the cell wall composition in the mature seeds of parents (L4602, and L830), 10 extreme large-seeded RILs (No. 39, 86, 87, 97, 102, 107, 108, 115, 133, 190) and 10 extreme small-seeded RILs (No. 5, 14, 16, 64, 88, 117, 155, 160, 168, 169) (Table 3). Nearly similar FT-IR cellulose content was recorded in the parental genotypes viz., L4602 (24.07%) and L830 (25.96%), while in large-seeded RILs, FT-IR cellulose content was recorded relatively less (21.25 to 28.63%) than that of the small-seeded RILs (22.42 to 39.16%). Lignin is a phenolic polymer that gives rigidity to cell wall, and FT-IR lignin content was recorded more in the small-seeded parental genotype L830 (12.80%) than the large-seeded parental genotype L4602 (11.16%). Similar observations were also recorded for the small-seeded RILs which showed relatively more lignin content (10.73 to 26.85%) than the large-seeded RILs (11.16 to 15.4%). Acetyl bromide soluble lignin (ABSL) content was recorded more in the small-seeded parental genotype L830 (25.07%) than the large-seeded parental genotype L4602 (21.98%). Similarly, small-seeded RILs showed more ABSL content (1.256 to 4.546%) than the large-seeded RILs (1.082 to 2.07%).

The xylan was recorded less than the cellulose or lignin in the seeds of parental genotypes and was found more in the small-seeded genotype L830 (6.86 mg/g) than the large-seeded genotype L4602 (4.16 mg/g). Similarly, in large-seeded RILs, xylan content ranged from 2.219 to 7.152 mg/g while in small-seeded RILs, it varied from 3.08 to 12.18 mg/g. Acetyl content was recorded more in small seeded genotype L830 (4.02 mg/g) than that of the large seeded genotype L4602 (2.013 mg/g). Similarly, in large-seeded RILs, acetyl content ranged from 2.013 to 6.138 mg/g; while in small-seeded RILs, it varied from 4.02 to 10.232 mg/g (Table 3). In general, cellulose was recorded as the most abundant cell wall component in lentil seeds. Overall, a higher value for almost all the studied cell wall components was recorded for the small-seeded genotype L830 over the large-seeded genotype L4602.

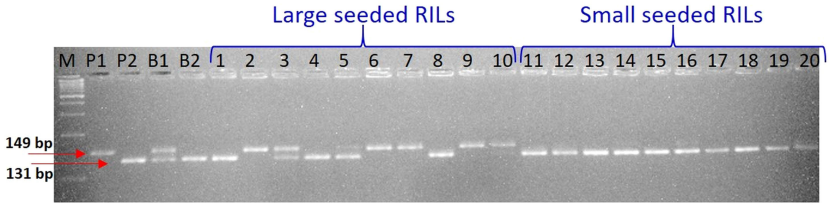


FIGURE 5
Gel picture showing individual plant analysis (constituting the bulks) for seed size with PBALC449 marker. Where, P1: L4602, P2: L830, B1: Large seeded bulk, B2: Small seeded bulk, 1-10: Individual plants constituting the large-seeded bulk; 11-20: Individual plants constituting the small seeded bulk, M: DNA ladder.

TABLE 2 The details of the seed size and the amplification pattern of the PBALC449 marker in the 188 RILs.

Band type	Band size (bp)	≤19.7g/1000 seed	>19.7-21g/1000 seed	21-23g/1000 seed	24-30g/1000 seed	>30g/1000 seed	Total
P1 (L4602, Large Seeded)	149	1	2	3	22	11	39
H (Heterozygous)	149 and 131	3	2	8	15	15	43
P2 (L830; Small Seeded)	131	33	25	16	25	7	106
Total	–	37	29	27	62	33	188
		93 (Small Seeded; 06 recombinant and 13 heterozygous)			95 (Large seeded; 32 recombinants and 30 heterozygous)		

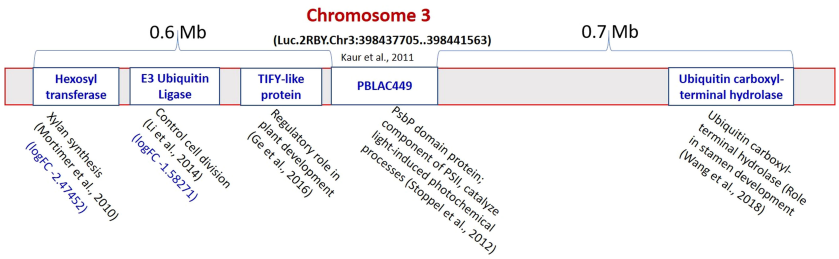


FIGURE 6
Putative genes identified regulating seed size on both sides (1.2 Mb region) of the PBALC449 in the lentil genome. (Derived from Mortimer et al., 2010; Stoppel et al., 2012; Li and Li, 2014; Ge et al., 2016; Wang et al., 2018).

Characterization of parental genotypes and the RILs using VideometerLab 4.0 for various seed parameters

The lentil parental genotypes L4602 (42.13 g/1000 seeds), L830 (20.90 g/1000 seeds) and the 10 extreme RILs (large seeded RILs: 34.7–39.2 g/1000 seed, and small-seeded RILs: 16.16–20.1 g/1000 seeds) differed significantly for the mean 1000 seed weight, were used for the study. Various other seed parameters like area, length, width, width/length, compactness, width/area, volume, and perimeter were also measured using VideometerLab 4.0 instrument, which showed significant variations for the studied genotypes (Table 4). Image of the lentil genotypes (L830 and L4602) as captured by VideometerLab 4.0 at 19 different wavelengths (375 to 970 nm) for further seed parameter analysis is given in Figure S8. Interestingly, the mean seed area (mm^2), length (mm), width (mm), and perimeter (mm) of the parental genotypes L4602 (22.59, 5.57, 5.24, 15.47 respectively) and L830 (11.02, 3.82, 3.71, 10.66 respectively) were found significantly different between these genotypes (Table 4). In addition, the 10 RILs (each extreme for seed size) were also compared through one-way ANOVA and were grouped using the Tukey HSD method ($p \leq 0.05$). For large seeded RILs, the studied seed parameters (like area: 17.36–20.82 mm^2, length: 4.9–5.3 mm, width: 4.56–5.05 mm, perimeter: 13.47–14.75 mm) were found significantly higher than the small seeded RILs (area: 8.88–11.03 mm^2, length: 3.51–3.83 mm, width: 3.28–3.71 mm, perimeter: 9.78–10.7 mm). ANOVA was also performed for all the 188 RILs (including parents) and details are presented in Table S4. A representative image (Figure S7) shows the details of four large and four small-seeded lentil RIL genotypes as captured by VideometerLab 4.0 at two wavelengths (590 and 850 nm).

10.3389/fpls.2023.1091432

TABLE 3 Estimation of cellulose, lignin, acetyl bromide soluble lignin content (ABSL), D- xylose, and acetylated xylose in the mature lentil seeds of 20 RILs (including both parents) and multiple comparison test using Tukey HSD method.

Parameters (Mean ± SD)	FT-IR cellulose (%)	FT-IR lignin (%)	Xylose content (mg/g)	O-Acetyl content (mg/g)	ABSL lignin content (%)
Large seeded					
L4602	24.07 ± 0.714bcdef	11.16 ± 0.049hi	4.163 ± 0.089efghi	**2.013 ± 0.166e**	2.197 ± 0.579cde
RIL 102	21.7 ± 2.263ef	12.99 ± 0.332defgi	**2.219 ± 0.170i**	5.425 ± 0.149bcd	2.070 ± 0.061de
RIL 097	26.1 ± 1.336bcdef	14.63± 0.183def	3.703 ± 0.578fghi	3.696 ± 0.513cde	1.898 ± 0.265de
RIL 039	28.23 ± 0.834bcd	15.15 ± 0.071de	5.401 ± 0.415efghi	3.912 ± 0.273bcde	1.873 ± 0.237de
RIL 107	23.19 ± 0.77cdef	12.15 ± 0.071fghi	3.518 ± 0.108ghi	3.292 ± 0.649de	1.082 ± 0.269e
RIL 190	28.2 ± 0.134bcd	14.63 ± 0.148def	6.492 ± 0.617defgh	5.074 ± 0.363bcd	1.423 ± 0.060e
RIL 133	28.5 ± 1.499bc	14.79 ± 0.82def	6.301 ± 0.634defghi	5.205 ± 0.108bcd	1.425 ± 0.018e
RIL 087	24.8 ± 0.424bcdef	13.75 ± 0.212defgh	6.622 ± 0.362cdefgh	4.290 ± 0.150bcde	1.516 ± 0.639e
RIL 108	28.63 ± 0.395bc	14.37 ± 0.198defg	6.618 ± 0.059cdefgh	5.055 ± 0.431bcd	1.461 ± 0.050e
RIL 086	26.81 ± 0.692bcdef	14.27 ± 0.099defg	7.152 ± 0.267cdefgh	6.138 ± 0.366bc	1.492 ± 0.171e
RIL 115	**21.25 ± 1.018bcd**	15.4 ± 0.989d	4.442 ± 0.053efghi	4.235 ± 0.389bcde	1.879 ± 0.308de
Small seeded					
L830	25.96 ± 0.064bcdef	12.8 ± 0.141defghi	6.86 ± 1.59cdefgh	4.02 ± 1.77bcde	2.506 ± 0.599bcde
RIL 014	29.93 ± 5.487b	13.28 ± 2.347defghi	7.70 ± 2.09bcdef	5.631 ± 0.0158 bcd	1.749 ± 0.300de
RIL 005	25.54 ± 0.346bcdef	11.35 ± 0.1767hi	9.598 ± 0.511abcd	6.514 ± 0.336b	1.456 ± 0.014e
RIL 064	27.68 ± 0.169bcde	12.28 ± 0.1697efghi	3.755 ± 0.376fghi	4.242 ± 0.190bcde	**1.256 ± 0.170e**
RIL 155	28.28 ± 0.049bcd	12.73 ± 0.587defghi	3.084 ± 0.273hi	4.067 ± 0.513bcde	2.475 ± 0.098bcde
RIL 160	25.59 ± 0.078bcdef	11.64 ± 0.318ghi	8.08 ± 1.52abcde	5.402 ± 0.359bcd	1.603 ± 0.250de
RIL 016	22.42 ± 1.407def	**10.73 ± 0.191i**	6.694 ± 0.980cdefgh	5.773 ± 0.344bcd	1.685 ± 0.066de
RIL 088	36.08 ± 0.898a	18.29 ± 0.035c	**12.180 ± 0.120a**	**10.232 ± 0.176a**	3.261 ± 0.692abcd
RIL 168	29.92 ± 0.7b	22.82 ± 0.863b	10.63 ± 2.13abc	6.324 ± 0.872bc	3.948 ± 0.904ab
RIL 169	29.44 ± 1.598b	**26.85 ± 0.629a**	7.19 ± 2.33cdefg	4.831 ± 0.367bcd	**4.546 ± 0.441a**
RIL 117	**39.16 ± 0.382a**	22.27 ± 1.421b	11.568 ± 0.544ab	9.80 ± 1.920a	3.835 ± 0.767abc

Values represent mean ± SD at P ≤ 0.05 and the same lower-case letters within a column are not significantly different. The values in bold represent the higher and lower values.

Values represent mean ± SD at P ≤ 0.05. Same lower-case letters within a column are not significantly different. The values in bold represent the higher and lower values.Validation of identified candidate genes in a RIL mapping population

Another mapping population (RIL; $F_{3:4}$) which was derived from the cross between Globe mutant (1000 seed weight=13.6g) and L4775 (1000 seed weight=28.47g) was used for the validation. Two contrasting bulks using 20 extreme plants each for the seed weight (small seed bulk: 1000 seed weight=18.57g; bold seed bulk: 1000 seed weight=24.46g) along with a parent (Globe Mutant) was used for the whole genome resequencing (WGRS). Detailed sequence analysis could identify 90 SNPs/InDels for the four candidate genes as identified by the BLAC449 marker (Table S5).

For *E3 ubiqutin ligase* gene 03 SNPs was identified; whereas for *TIFY-like protein* gene, 34 SNPs and 01 InDel was identified and most of these showed modifier effect. Among the 20 SNPs and 02 InDels of *Hexosyltransferase* gene, one InDel showed disruptive inframe deletion with moderate effect while other SNPs showed mostly missense variant with moderate effect. Similarly, for the *Ubiquitin carboxyl-terminal hydrolase* gene we have identified 30 SNPs and most of these showed modifier effect (Table S5).

Discussion

BSA and identification of candidate genes regulating seed size trait in lentil

A total of 394 SSR diverse SSR primer pairs (Saha et al., 2010; Kaur et al., 2011; Jain et al., 2013; Verma et al., 2014) were used and 31 were found polymorphic, which is 7.9% of the total primers used. A similar level of polymorphism was also reported by previous workers (Kumar et al., 2019; Singh et al., 2019). Of all the polymorphic SSRs,

10.3389/fpls.2023.1091432

TABLE 4 List of seed parameters studied using VideometerLab 4.0 for parents (L4602 and L830) and the 10 extreme RILs. .

Parameters (Mean ± SD)									
Genotypes	1000 seed wt (g)	Area (mm^2)	Length (mm)	Width (mm)	Width/ length	Compactness (Circle)	Width/ area	Volume (mm^3)	Perimeter (mm)
Parents									
L4602	42.13 ± 1.21a	22.59 ± 1.17a	5.57 ± 0.169a	5.24 ± 0.179a	0.939 ± 0.027cde	0.938 ± 0.025bcd	0.232 ± 0.006j	0.007 ± 0.0002a	15.47 ± 0.452a
L830	20.90 ± 1.82d	11.02 ± 0.68g	3.82 ± 0.121gh	3.71 ± 0.117g	0.972 ± 0.015a	0.969 ± 0.014a	0.338 ± 0.011e	0.005 ± 0.0001gh	10.66 ± 0.369h
Large seeded RILs (10 No)									
RIL039	**39.20 ± 1.368ab**	20.24 ± 0.633bc	5.24 ± 0.08b	4.94 ± 0.123bcd	**0.942 ± 0.019bcde**	0.941 ± 0.016bcd	0.244 ± 0.003i	0.0065 ± 0.0001b	14.60 ± 0.23bc
RIL086	37.43 ± 2.71bc	**17.36 ± 1.396f**	**4.90 ± 0.156f**	**4.56 ± 0.199f**	0.961 ± 0.022e	**0.927 ± 0.037d**	**0.340 ± 0.009f**	**0.00616 ± 0.00019f**	**13.47 ± 0.463g**
RIL087	**34.70 ± 0.608c**	18.30 ± 0.889e	4.95 ± 0.119ef	4.77 ± 0.148e	0.964 ± 0.019abc	0.961 ± 0.017ab	0.261 ± 0.006fg	0.006223 ± 0.00015ef	13.77 ± 0.332fg
RIL097	37.70 ± 0.608bc	**20.82 ± 0.737b**	**5.30 ± 0.114b**	**5.05 ± 0.117b**	0.952 ± 0.025abcde	0.952 ± 0.022abcd	**0.242 ± 0.006i**	**0.006668 ± 0.00014b**	**14.75 ± 0.266b**
RIL102	39.16 ± 1.607ab	18.73 ± 0.614e	5.06 ± 0.113cde	4.85 ± 0.147de	0.958 ± 0.035abcd	0.955 ± 0.028abc	0.259 ± 0.007fgh	0.006366 ± 0.00014cde	14.27 ± 0.391cde
RIL107	38.46 ± 0.924abc	19.57 ± 0.74cd	5.17 ± 0.124bc	4.9 ± 0.116cde	0.948 ± 0.022abcde	0.946 ± 0.023abcd	0.250 ± 0.005ghi	0.006507 ± 0.00016bc	14.27 ± 0.25cde
RIL108	35.93 ± 1.102bc	20.09 ± 0.658bc	5.19 ± 0.123bc	5.03 ± 0.096bc	**0.968 ± 0.016ab**	**0.968 ± 0.014a**	0.250 ± 0.005hi	0.006527 ± 0.00015bc	14.43 ± 0.272bcd
RIL115	35.73 ± 1.419bc	18.92 ± 0.902de	5.09 ± 0.117cd	4.83 ± 0.169de	0.949 ± 0.029abcde	0.947 ± 0.028abcd	0.255 ± 0.006fgh	0.006396 ± 0.00015cd	14.21 ± 0.299de
RIL133	35.56 ± 0.513bc	20.63 ± 0.608b	5.28 ± 0.086b	5.04 ± 0.088bc	0.954 ± 0.018abcde	0.953 ± 0.018abcd	0.244 ± 0.004i	0.006643 ± 0.00011bde	14.67 ± 0.269b
RIL190	36.43 ± 0.512bc	18.83 ± 0.794de	5.05 ± 0.121de	4.83 ± 0.104de	0.957 ± 0.015abcde	0.959 ± 0.016ab	0.256 ± 0.006fgh	0.006344 ± 0.00015de	13.98 ± 0.349ef
Small seeded RILs (10 No)									
RIL005	**20.1 ± 1.85de**	**11.03 ± 0.462 g**	**3.83 ± 0.076g**	**3.71 ± 0.112g**	**0.967 ± 0.016abc**	0.968 ± 0.014a	**0.336 ± 0.0062e**	**0.00482 ± 0.000096g**	**10.7 ± 0.23h**
RIL014	18.13 ± 1.026de	10.57 ± 0.466gh	3.79 ± 0.09ghi	3.60 ± 0.089gh	0.947 ± 0.026abcde	0.948 ± 0.027abcd	0.341 ± 0.0113de	0.004768 ± 0.000113ghi	10.4 ± 0.224hi
RIL016	17.03 ± 1.002de	10.67 ± 0.438gh	3.79 ± 0.088ghi	3.64 ± 0.098gh	0.960 ± 0.023abcd	0.960 ± 0.023ab	0.341 ± 0.0085de	0.004763 ± 0.000112ghi	10.52 ± 0.222hi
RIL064	17.83 ± 1.041de	9.57 ± 0.309ij	3.60 ± 0.079Jkl	3.42 ± 0.076ij	0.950 ± 0.027abcde	0.952 ± 0.022abcd	0.357 ± 0.0097bc	0.004526 ± 0.0001jkl	10.01 ± 0.184jkl
RIL088	17.3 ± 0.608de	**8.88 ± 0.481j**	**3.51 ± 0.095l**	**3.28 ± 0.113k**	**0.934 ± 0.031de**	**0.93 ± 0.033cd**	**0.369 ± 0.0128a**	**0.004413 ± 0.000121l**	9.79 ± 0.319l
RIL111	18.6 ± 1.637de	9.16 ± 0.355j	3.55 ± 0.085kl	3.36 ± 0.079jk	0.940 ± 0.029abcde	0.938 ± 0.028bcd	0.367 ± 0.011ab	0.004469 ± 0.000107kl	9.89 ± 0.263kl
RIL117	16.66 ± 1.528e	9.23 ± 0.483j	3.54 ± 0.124kl	3.37 ± 0.073jk	0.951 ± 0.028abcde	0.951 ± 0.023abcd	0.365 ± 0.0137ab	0.004456 ± 0.000156kl	**9.78 ± 0.292l**
RIL155	17.13 ± 1.026de	10.09 ± 0.768hi	3.66 ± 0.129ijk	3.55 ± 0.146hi	0.9672 ± 0.021ab	**0.97 ± 0.017a**	0.352 ± 0.0146c	0.004609 ± 0.000162ijk	10.23 ± 0.405ijk
RIL160	18.8 ± 1.311de	10.13 ± 0.449hi	3.69 ± 0.085hij	3.56 ± 0.083hi	0.963 ± 0.015abc	0.962 ± 0.015ab	0.351 ± 0.008c	0.004647 ± 0.000108hij	10.23 ± 0.259ijk
RIL169	**16.16 ± 0.289e**	10.3 ± 0.431ghi	3.73 ± 0.079ghi	3.58 ± 0.072gh	0.960 ± 0.014abcd	0.962 ± 0.015ab	0.348 ± 0.0091cd	0.004692 ± 0.0001ghi	10.35 ± 0.225hij

only one (PBLAC449) could differentiate the small seed size bulk and the parent, while the large seed size bulk showed two bands. This kind of unique polymorphism pattern was not yet reported in the lentil. Detailed RIL analysis (188 No) using PBLAC449 marker showed that the region near the PBLAC449 marker, seems to regulate the small seed size trail while large seed size is being governed by more than one locus. Moreover, quantitative regulation of seed size trait is reported by a number of workers (Fedoruk et al., 2013; Verma et al., 2015; Khazaei et al., 2018).

To understand this unique type of banding pattern, and to find the chromosomal location of amplified product (tightly linked with the small seed size trait only); cloning, sequencing, and the comparative genomics approaches were used. The PCR amplified fragment was cloned, sequenced, and aligned to the recently released lentil reference genome (CDC Redberry Genome Assembly v2.0) (Ramsay et al., 2021). The PCR amplified product from PBALC449 got aligned at Luc.2RBY.Chr3:398437705.398441563 (+strand) on chromosome number 3 and is a PsbP domain protein-encoding gene (3859 bp) (Figure S5). Similarly, Verma et al. (2015) have identified three major QTLs for seed weight and seed size traits in lentils on LG4; while, Fedoruk et al. (2013) have identified three QTLs for seed diameter on LG1, 2, and 7 which together explained >60% of the PVE and Khazaei et al. (2018) have identified two associated SNPs with seed diameter (*viz*. LcC09638p190 and LcC08798p992) on chromosomes 1 and 7, respectively. In addition, QTLs for seed weight (Abbo et al., 1991) and seed diameter (Fratini et al., 2007) are identified in lentils.

Further, to identify the candidate genes regulating small seed size trait near this marker (1.4 Mb region), we analyzed our RNA Seq data (Dutta et al., 2022). On the left side of the PBALC449 amplified region (0.6 Mb), three genes namely, E3 ubiquitin ligase (log2FC -1.582), hexosyltransferase (log2FC -2.474), and TIFY-like protein gene were found, while on the right side (0.7 Mb) a ubiquitin carboxyl-terminal hydrolase gene was found (Figure 6). The E3 Ubiquitin ligase gene is known to have a role in controlling cell division (Li and Li, 2014); while the TIFY-like protein gene is having a role in regulating the process of plant development (Ge et al., 2016). Similarly, the hexosyltransferase gene is known to have a role in the regulation of xylan synthesis (Mortimer et al., 2010); while ubiquitin carboxyl-terminal hydrolase gene is required for periodic maintenance of the circadian clock (Hayama et al., 2019) and inflorescence architecture (Yang et al., 2007) in *Arabidopsis*.

Domoney et al. (2006) reported two distinct phases during seed development in the legumes. In the first phase, cell division (in seeds) is dependent on embryo genotype having certain loci controlling the cotyledon cell number and is largely insensitive to environmental cues. Thus, this phase mainly controls the seed diameter and seed plumpness. The second phase regulates seed thickness *via* cell expansion, which is highly influenced by the environment and is mainly regulated by photosynthate partitioning loci. The seed size is reportedly influenced by both pre-anthesis and post-anthesis periods (Gupta et al., 2006) as these affect the amount of assimilates partitioned to the developing seeds (Pre-anthesis) and also the time for seed maturation (post-anthesis) which could alter the seed size. Flowering time and other flower morphology-related loci were also known to control the seed size in model legume crops (Ohto et al., 2005; He et al., 2010; Wang et al., 2012). In chickpea, a major flowering time gene, PPD, is reported to affect the seed weight, and early flowering results in reduced seed weight (Hovav et al., 2003). Validation results in another mapping population (Globe mutant × L4775) using WGRS also confirmed the presence of SNPs and InDels in the four candidate genes. However, there is still a need to validate these candidate genes having a role in the seed size regulation, in different lentil genotypes for its ultimate application in the breeding program aiming for seed size improvement.

Seed biochemical parameters

Seed size and shape are regulated by the cell wall composition in lentils (Dutta et al., 2022). However, no other detailed report mentioning the relationship between the seed size and cell wall composition including cellulose, lignin, and xylose in lentils is known. The data of parents and the 10 extreme RILs for the cell wall composition in the mature seeds showed significant variations for parameters like FT-IR cellulose, FT-IR lignin, ABSL, xylan, and acetyl content (Table 3). In small seeded RILs, in general, more of FT-IR cellulose (22.4 to 39.16%), FT-IR lignin (10.73 to 26.85%), ABSL (1.26 to 4.55%), xylose content (3.083 to 12.18 mg/g), acetyl content (4.02 to 10.23 mg/g) was recorded than the large-seeded RILs (FT-IR cellulose: 21.25 to 28.6%; FT-IR lignin: 11.16 to 15.4%; ABSL: 1.08 to 2.19%; xylose content: 2.22 to 7.15 mg/g; acetyl content: 2.01 to 6.14 mg/g). Overall, cellulose was recorded as the most abundant cell wall component in lentil seeds. Similarly, cellulose and hemicellulose such as galactomannan, mannan, and xyloglucan were found to play a crucial function in determining the shape and size of both developing and mature seeds (Buckeridge, 2010).

This study recorded up to 39.16% cellulose (FT-IR) in lentil seeds, whereas in different plant species nearly 40–60% cellulose was reported (Costa and Plazanet, 2016). In the RILs, 10.73 to 26.85% lignin (FT-IR) was recorded whereas 5.13% mean lignin content was recorded in soybean seeds (Krzyzanowski et al., 2008), and genotypes having >5% lignin in the seed coat were less prone to mechanical damage (Alvarez et al., 1997). The presence of more lignin in lentil seeds over soybean may be due to the presence of more colored compounds in the lentil seed coat (Dutta et al., 2022). Xylose and xyloglucan are considered important seed storage polysaccharides, especially in developing seeds (Buckeridge, 2010). The studied lentil genotypes showed 2.22 to 12.18 mg/g xylose content, whereas 3.5–4.5% (dry weight basis) acetyl–xylose content was recorded in the hardwoods (Teleman et al., 2002). Acetyl content in the range of 2.01 to 10.23 mg/g was recorded in the studied RILs. Differential seed sizes in different genotypes might be due to the different levels of polysaccharides acetylation which seems to affect their water solubility, interactions with cellulose, and various other physicochemical properties (Busse-Wicher et al., 2014).

RNA-seq results of Dutta et al. (2022) identified various cell wall-associated GO terms and also the differential expression of xyloglucan endotransglucosylase encoding gene, suggesting their involvement in the cell wall synthesis during seed development in lentils, and similar results were also recorded in soybean (Du et al., 2017). Overall, a higher value for almost all the studied cell wall components for small-seeded lentil genotype (L830) over large-seeded (L4602) genotype needs further detailed stage-specific investigations.

Characterization of lentil genotypes using VideometerLab 4.0 for various seed parameters

In general, seed size in lentils is mostly determined using a very crude method of measuring 100 or 1000 seed weight (Tullu et al., 2001). Even in soybean, seed shape parameters are measured using a caliper (Xu et al., 2011), while in chickpeas Hossain et al. (2010) used seed sizing using graded sieves for determining the seed size and shape. By this, it is impossible to determine the seed thickness or seed plumpness (Dutta et al., 2022). However, in this study, VideometerLab 4.0 instrument was used to measure various seed parameters like area, length, width, width/length, compactness, width/area, volume, and perimeter of all the RILs (188 No) and its parents. Most of the studied parameters showed significant variations for the studied genotypes (Table S4). For large seeded RILs and parents, 1000 seed weight (34.7 to 42.13 g), area (17.36 to 22.59 mm^2), seed length (4.9 to 5.57 mm), seed width (4.56 to 5.24 mm), and seed perimeter (13.47 to 15.47 mm) were found significantly more than the small seeded RILs including parent (1000 seed weight: 16.16 to 20.90 g, area: 8.88 to 11.03 mm^2, seed length: 3.51 to 3.83 mm, seed width: 3.28 to 3.71 mm, and seed perimeter: 9.78 to 10.7 mm). Similarly, Shahin et al. (2012) used cameras and captured the 3-dimensional lentil seed images and measured the seed plumpness; while Shahin and Symons (2001) deployed computer-aided two-dimensional imaging to measure the diameter of lentil seeds. Similarly, previous studies also demonstrated huge variations for various seed parameters in lentils (Tullu et al., 2001; Tripathi et al., 2022). Thus, the use of VideometerLab was found very precise, quick, and easy method for the determination of several seed parameters.

Conclusions

Results of the study have conclusively shown the importance of the maker PBLAC449 in the identification of genotypes having small seed size in lentils. In addition, the region identified on chromosome 03, needs more critical attention for the validation of genes regulating the seed size trait in lentils. The cell wall composition including cellulose, xylan, etc. was extensively analyzed, using wet chemistry methods and FT-IR to understand the association between cell extensibility and the seed size in lentils. Compared to any other method, the use of VideometerLab 4.0 was found very effective, easy, and quick, and should be used for the measurement of various essential seed parameters in lentils. Thus, the information generated in this study has paved the way for further in-depth analysis of the factors governing seed size in lentils including the development of genotypes having customized seed sizes.

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

Conceptualization: GM, HKD, SG, and SK; **methodology**: HD, SM, SS, MA, MT, SD, PP, AK, KT, and RK; **formal analysis:** HD, SM, DV, DM, and AS; **resources**: GM, SK, and HKD; **data curation**: HKD and GM; **writing—original draft preparation**: HD, GM, and HKD; **writing—review and editing**: HKD, GM, SG, and SK; **supervision**: GM and HKD. All authors contributed to the article and approved the submitted version.

Funding

The work was supported and funded by the Indian Council of Agricultural Research (ICAR) and the International Center for Agricultural Research in the Dry Areas (ICARDA).

Acknowledgments

Technical support received from Mr. Dilip Kumar and the access to VideometerLab 4.0 instrument given by Mr. Bharatkumar Mathuradas Davda, Founder, Tara International (Vadodara, Gujarat) at ICAR-IARI, New Delhi is duly acknowledged.

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1091432/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Seed size variation among parents (L4602 and L830) and the 10 extreme RILs for large and small seed size.

SUPPLEMENTARY FIGURE 2
A representative gel pictures showing results of parental polymorphism between L4602 and L830 for various SSR markers. Where 1. is L4602 and 2. is L830

SUPPLEMENTARY FIGURE 3
BLAST for sequence from large seeded parent (L4602) (149 bp).

SUPPLEMENTARY FIGURE 4
BLAST for sequence from small seeded parent (L830) (131 bp; due to 18 bp deletion at two places).

SUPPLEMENTARY FIGURE 5
Identified position of the SSR (PBALC 449) in the lentil genome (Chromosome 3).

SUPPLEMENTARY FIGURE 6
Sequnce similarity of the marker (PBALC449) with the relative sp. It is showing most similarity to *Medicago truncatula*, followed by *Cicer arietinum*.

SUPPLEMENTARY FIGURE 7
Representative image of eight lentil RIL genotypes, captured by VideometerLab 4.0 at two wavelengths (590 and 850 nm) for further seed parameter analysis.

SUPPLEMENTARY FIGURE 8
Image of the lentil genotypes (L830 and L4602) captured by VideometerLab 4.0 at 19 wavelengths (375 to 970 nm) for further seed parameter analysis.

References

Abbo, S., Ladizinsky, G., and Weeden, N. F. (1991). Genetic-analysis and linkage study of seed weight in lentil. *Euphytica* 58, 259–266. doi: 10.1007/BF00025258

Alvarez, P. J. C., Krzyzanowski, F. C., Mandarino, J. M. G., and França Neto, J. B. (1997). Relationship between soybean seed coat lignin content and resistance to mechanical damage. *Seed. Sci. Technol.* 25, 209–214.

Arumuganathan, K., and Earle, E. D. (1991). Nuclear DNA content of some important plant species. *Plant Mol. Biol. Rep.* 9, 208–218. doi: 10.1007/BF02672069

Barulina, H. (1930). Lentils of the USSR and other countries (English summary). *Bull. Appl. Bot. Genet. Plant Breed.* 40, 265–304.

Buckeridge, M. S. (2010). Seed cell wall storage polysaccharides: Models to understand cell wall biosynthesis and degradation. *Plant Physiol.* 154 (3), 1017–1023. doi: 10.1104/pp.110.158642

Busse-Wicher, M., Gomes, T. C., Tryfona, T., Nikolovski, N., Stott, K., Grantham, N. J., et al. (2014). The pattern of xylan acetylation suggests xylan may interact with cellulose microfibrils as a twofold helical screw in the secondary plant cell wall of *Arabidopsis thaliana*. *Plant J.* 79 (3), 492–506. doi: 10.1111/tpj.12575

Canteri, M., Renard, C., Le Bourvellec, C., and Bureau, S. (2019). ATR-FTIR spectroscopy to determine cell wall composition: Application on a large diversity of fruits and vegetables. *Carbohydr. Polymers.* 212, 186–196. doi: 10.1016/j.carbpol.2019.02.021

Costa, G., and Plazanet, I. (2016). Plant cell wall, a challenge for its characterisation. *Adv. Biol. Chem.* 6, 70–105. doi: 10.4236/abc.2016.63008

Dikshit, H. K., Mishra, G. P., Aski, M. S., Singh, A., Tripathi, K., Bansal, R., et al. (2022a). "Lentil breeding," in *Fundamentals of field crop breeding*. Eds. D. K. Yadava, H. K. Dikshit, G. P. Mishra and S. Tripathi (Singapore: Springer), 1181–1236. doi: 10.1007/978-981-16-9257-4_24

Dikshit, H. K., Mishra, G. P., Aski, M., Singh, A., Virk, P. S., and Kumar, S. (2022b). "Lentil biofortification," in *Biofortification of staple crops*. Eds. S. Kumar, H. K. Dikshit, G. P. Mishra and A. Singh (Singapore: Springer), 271–293. doi: 10.1007/978-981-16-3280-8_11

Domoney, C., Duc, G., Ellis, T. H., Ferrándiz, C., Firnhaber, C., Gallardo, K., et al. (2006). Genetic and genomic analysis of legume flowers and seeds. *Curr. Opin. Plant Biol.* 9, 133–141. doi: 10.1016/j.pbi.2006.01.014

Du, J., Wang, S., He, C., Zhou, B., Ruan, Y. L., Shou, H., et al. (2017). Identification of regulatory networks and hub genes controlling soybean seed set and size using RNA sequencing analysis. *J. Exp. Bot.* 68 (8), 1955–1972. doi: 10.1093/jxb/erw460

Dutta, H., Mishra, G. P., Aski, M. S., Bosamia, T. C., Mishra, D. C., Bhati, J., et al. (2022). Comparative transcriptome analysis, unfolding the pathways regulating the seed-size trait in cultivated lentil (*Lens culinaris* medik.). *Front. Genet.* 13. doi: 10.3389/fgene.2022.942079

Erskine, W., Williams, P. C., and Nakkoul, H. (1991). Splitting and dehulling lentil (*Lens culinaris*): Effects of seed size and different pretreatments. *J. Sci. Food Ag.* 57, 77–84. doi: 10.1002/jsfa.2740570109

FAOSTAT (2020) *Statistical databases* (Italy: Food and Agriculture Organization of the United Nations). Available at: https://www.fao.org/faostat/en/#home (Accessed Aug 14, 2022).

Fedoruk, M., Vandenberg, A., and Bett, K. (2013). Quantitative trait loci analysis of seed quality characteristics in lentil using single nucleotide polymorphism markers. *Plant Genome* 6 (3), 1–10. doi: 10.3835/plantgenome2013.05.0012

Ford, R., Rubeena Redden, R. J., Materne, M., and Taylor, P. W. J.. (2007). Lenti. In: C. Kole eds. *Genome Mapping and Molecular Breeding in Plants* (Berlin, Heidelberg: Springer) 3, 91–108. doi: 10.1007/978-3-540-34516-9_5

Foster, C. E., Martin, T. M., and Pauly, M. (2010). Comprehensive compositional analysis of plant cell walls (lignocellulosic biomass) part I: lignin. *J. Visualized. Experiments.* 37, 1745. doi: 10.3791/1745

Fratini, R., Duran, Y., Garcia, P., and Perez de la Vega, M. (2007). Identification of quantitative trait loci (QTL) for plant structure, growth habit and yield in lentil. *Spanish. J. Ag. Res.* 5, 348–356. doi: 10.5424/sjar/2007053-255

García-Alegría, A., Anduro-Corona, I., Pérez-Martínez, C., Guadalupe Corella-Madueño, M., Rascón-Durán, M., and Astiazaran-Garcia, H. (2020). Quantification of DNA through the NanoDrop spectrophotometer: Methodological validation using standard reference material and sprague dawley rat and human DNA. *Int. J. Anal. Chem.* 2020, 1–9. doi: 10.1155/2020/8896738

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. Nat. Acad. Sci. U.S.A.* 113 (44), 12414–12419. doi: 10.1073/pnas.1611763113

Gupta, P. K., Rustgi, S., and Kumar, N.. (2006). Genetic and molecular basis of grain size and grain number and its relevance to grain productivity in higher plants. *Genome* 49 (6), 565–571. doi: 10.1139/g06-063

Hamdi, A., Erskine, W., and Gates, P.. (1991). Relationships among economic characters in lentil. *Euphytica* 57 (2), 109–116. doi: 10.1007/bf00023068

Hayama, R., Yang, P., Valverde, F., Mizoguchi, T., Furutani-Hayama, I., Vierstra, R. D., et al. (2019). Ubiquitin carboxyl-terminal hydrolases are required for period maintenance of the circadian clock at high temperature in *Arabidopsis*. *Sci. Rep.* 9, 17030. doi: 10.1038/s41598-019-53229-8

He, C., Tian, Y., Saedler, R., Efremova, N., Riss, S., Khan, M. R., et al. (2010). The MADS-domain protein MPF1 of *Physalis floridana* controls plant architecture, seed development and flowering time. *Planta* 231, 767–777. doi: 10.1007/s00425-009-1087-z

Hossain, S., Ford, R., McNeil, D., Pittock, C., and Panozzo, J. F. (2010). Development of a selection tool for seed shape and QTL analysis of seed shape with other morphological traits for selective breeding in chickpea (*Cicer arietinum* l.). *Aus. J. Crop Sci.* 4, 278–288.

Hovav, R., Upadhyaya, K. C., Beharav, A., and Abbo, S. (2003). Major flowering time gene and polygene effects on chickpea seed weight. *Plant Breed.* 122, 539–541. doi: 10.1111/j.1439-0523.2003.00895.x

Jain, N., Dikshit, H. K., Singh, D., Singh, A., and Kumar, H. (2013). Discovery of EST-derived microsatellite primers in the legume *Lens culinaris* (Fabaceae). *Appl. Pl. Sci.* 1 (7), 1200539. doi: 10.3732/apps.1200539

Kaur, S., Cogan, N., Pembleton, L., Shinozuka, M., Savin, K., Materne, M., et al. (2011). Transcriptome sequencing of lentil based on second-generation technology permits large-scale unigene assembly and SSR marker discovery. *BMC Genomics* 12 (1), 265. doi: 10.1186/1471-2164-12-265

Khazaei, H., Fedoruk, M., Caron, C., Vandenberg, A., and Bett, K. (2018). Single nucleotide polymorphism markers associated with seed quality characteristics of cultivated lentil. *Plant Genome* 11 (1), 170051. doi: 10.3835/plantgenome2017.06.0051

Krzyzanowski, F., Franca Neto, J., Mandarino, J., and Kaster, M. (2008). Evaluation of lignin content of soybean seed coat stored in a controlled environment. *Rev. Bras. Sementes.* 30 (2), 220–223. doi: 10.1590/s0101-31222008000200028

Kumar, H., Singh, A., Dikshit, H. K., Mishra, G. P., Aski, M., Meena, M. C., et al. (2019). Genetic dissection of grain iron and zinc concentrations in lentil (*Lens culinaris* medik.). *J. Genet.* 98 (3), 66. doi: 10.1007/s12041-019-1112-3

Labbe, N., Rials, T., Kelley, S., Cheng, Z., Kim, J., and Li, Y. (2005). FT-IR imaging and pyrolysis-molecular beam mass spectrometry: new tools to investigate wood tissues. *Wood Sci. Technol.* 39 (1), 61–76. doi: 10.1007/s00226-004-0274-0

Li, N., and Li, Y. (2014). Ubiquitin-mediated control of seed size in plants. *Front. Plant Sci.* 5. doi: 10.3389/fpls.2014.00332

Michelmore, R., Paran, I., and Kesseli, R. (1991). Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. *Proc. Nat. Acad. Sci. U.S.A.* 88 (21), 9828–9832. doi: 10.1073/pnas.88.21.9828

Mishra, G. P., Ankita,, Aski, M. S., Tontang, M. T., Priti, C., Tripathi, K., et al. (2022b). Morphological, molecular, and biochemical characterization of a unique lentil (*Lens culinaris* medik.) genotype showing seed-coat color anomalies due to altered anthocyanin pathway. *Plants* 11 (14), 1815. doi: 10.3390/plants11141815

Mishra, G. P., Aski, M. S., Bosamia, T., Chaurasia, S., Mishra, D. C., Bhati, J., et al. (2022a). Insights into the host-pathogen interaction pathways through RNA-seq analysis of *Lens culinaris* medik. in response to *Rhizoctonia bataticola* infection. *Genes* 13 (1), 90. doi: 10.3390/genes13010090

Mishra, G. P., Dikshit, H. K., Kumari, J., Priti,, Tripathi, K., Devi, J., et al. (2020). Identification and characterization of novel penta-podded genotypes in the cultivated lentil (*Lens culinaris* medik.). *Crop Sci.* 60 (4), 1974–1985. doi: 10.1002/csc2.20156

Mishra, G. P., Singh, R. K., Mohapatra, T., Singh, A. K., Prabhu, K. V., and Zaman, F. U. (2001). Molecular mapping of a fertility restorer gene in basmati rice using microsatellite markers. *Indian J. Genet. Plant Breed.* 61 (4), 348–349.

Mishra, G. P., Singh, R. K., Mohapatra, T., Singh, A. K., Prabhu, K. V., Zaman, F. U., et al. (2003). Molecular mapping of a gene for fertility restoration of wild abortive (WA) cytoplasmic male sterility using a basmati line restorer line. *J. Plant Biochem. Biotechnol.* 12, 37–42. doi: 10.1007/BF03263157

Mortimer, J., Miles, G., Brown, D., Zhang, Z., Segura, M., Weimar, T., et al. (2010). Absence of branches from xylan in *Arabidopsis* gux mutants reveals potential for simplification of lignocellulosic biomass. *Proc. Nat. Acad. Sci. U.S.A.* 107 (40), 17409–17414. doi: 10.1073/pnas.1005456107

Murray, M., and Thompson, W. (1980). Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res.* 8 (19), 4321–4326. doi: 10.1093/nar/8.19.4321

Ohto, M. A., Fischer, R. L., Goldberg, R. B., Nakamura, K., and Harada, J. J. (2005). Control of seed mass by *APETALA2*. *Proc. Nat. Acad. Sci. U.S.A.* 102, 3123–3128. doi: 10.1073/pnas.0409858102

Pawar, P., Derba-Maceluch, M., Chong, S., Gandla, M., Bashar, S., Sparrman, T., et al. (2017). In muro deacetylation of xylan affects lignin properties and improves saccharification of aspen wood. *Biotechnol. Biofuels* 10 (1), 98. doi: 10.1186/s13068-017-0782-4

Priti., Mishra, G. P., Dikshit, H. K., Vinutha, T., Mechiya, T., Stobdan, T., et al. (2021). Diversity in phytochemical composition, antioxidant capacities, and nutrient contents among mungbean and lentil microgreens when grown at plain-altitude region (Delhi) and high-altitude region (Leh-ladakh), India. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.710812

Priti., Sangwan, S., Kukreja, B., Mishra, G. P., Dikshit, H. K., Singh, A., et al. (2022). Yield optimization, microbial load analysis, and sensory evaluation of mungbean (*Vigna radiata* l.), lentil (*Lens culinaris* subsp. *culinaris*), and Indian mustard (*Brassica juncea* l.) microgreens grown under greenhouse conditions. *PloS One* 17, e0268085. doi: 10.1371/journal.pone.0268085

Ramsay, L., Koh, C. S., Kagale, S., Gao, D., Kaur, S., Haile, T., et al. (2021) *Genomic rearrangements have consequences for introgression breeding as revealed by genome assemblies of wild and cultivated lentil species. bioRxiv. 2021 jul 24*. Available at: https://knowpulse.usask.ca/genome-assembly/Lcu.2RBY.

Rastogi, L., Chaudhari, A., Sharma, R., and Pawar, P. (2022). Arabidopsis GELP7 functions as a plasma membrane-localized acetyl xylan esterase, and its overexpression improves saccharification efficiency. *Plant Mol. Biol.* 109 (6), 781–797. doi: 10.1007/s11103-022-01275-8

Saha, G., Sarker, A., Chen, W., Vandemark, G., and Muehlbauer, F. (2010). Inheritance and linkage map positions of genes conferring resistance to *Stemphylium* blight in lentil. *Crop Sci.* 50 (5), 1831–1839. doi: 10.2135/cropsci2009.12.0709

Sambrook, J., Fritsch, E. R., and Maniatis, T. (1989). *Molecular cloning: A laboratory manual. 2nd ed* (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press), 1546, ISBN: .

Sandhu, J., and Singh, S. (2007). History and origin. In: Eds. S. S. Yadav, D. L. McNeil and P. C. Stevenson *Lentil*. (Dordrecht: Springer). doi: 10.1007/978-1-4020-6313-8_1

Shahin, M. A., and Symons, S. J. (2001). A machine vision system for grading lentils. *Cand. Biosys. Eng.* 43, 7–7.

Shahin, M. A., Symons, S. J., and Poysa, V. W. (2006). Determining soya bean size uniformity using image analysis. *Biosys. Eng.* 94, 191–198. doi: 10.1016/j.biosystemseng.2006.02.011

Shahin, M. A., Symons, S. J., and Wang, N. (2012). Predicting dehulling efficiency of lentils based on seed size and shape characteristics measured with image analysis. *Qual. Assur. Saf. Crops Foods.* 4, 9–16. doi: 10.1111/j.1757-837X.2011.00119.x

Shrestha, S., Deleuran, L., Olesen, M., and Gislum, R. (2015). Use of multispectral imaging in varietal identification of tomato. *Sensors* 15 (2), 4496–4512. doi: 10.3390/s150204496

Singh, A., Dikshit, H. K., Mishra, G. P., Aski, M., and Kumar, S. (2019). Association mapping for grain diameter and weight in lentil using SSR markers. *Plant Gene* 20, 100204. doi: 10.1016/j.plgene.2019.100204

Singh, A., Dikshit, H. K., Mishra, G. P., Aski, M., Kumar, S., and Sarker, A. (2022). Breeding for abiotic stress tolerance in lentil in genomic era. In *Genomic designing for abiotic stress resistant pulse crops*. Ed. C. Kole (Cham: Springer), 145. doi: 10.1007/978-3-030-91039-6_5

Sonnante, G., Hammer, K., and Pignone, D. (2009). From the cradle of agriculture a handful of lentils: history of domestication. *Rendiconti. Lincei.* 20, 21–37. doi: 10.1007/s12210-009-0002-7

Stoppel, R., Manavski, N., Schein, A., Schuster, G., Teubner, M., Schmitz-Linneweber, C., et al. (2012). RHON1 is a novel ribonucleic acid-binding protein that supports RNase e function in the *Arabidopsis* chloroplast. *Nucleic Acids Res.* 40, 8593–8606. doi: 10.1093/nar/gks613

Teleman, A., Tenkanen, M., Jacobs, A., and Dahlman, O. (2002). Characterization of O-acetyl-(4-O-methylglucurono) xylan isolated from birch and beech. *Carbohydr. Res.* 337 (4), 373–377. doi: 10.1016/s0008-6215(01)00327-5

Tripathi, K., Kumari, J., Gore, P. G., Mishra, D. C., Singh, A. K., Mishra, G. P., et al. (2022). Agro-morphological characterization of lentil germplasm of Indian national genebank and development of a core set for efficient utilization in lentil improvement programs. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.751429

Tullu, A., Kusmenoglu, I., McPhee, K. E., and Muehlbauer, F. J. (2001). Characterization of core collection of lentil germplasm for phenology, morphology, seed and straw yields. *Gen. Res. Crop Evol.* 48, 143–152. doi: 10.1023/A:1011254629628

Updegraff, D. M. (1969). Semimicro determination of cellulose in biological materials. *Anal. Biochem.* 32 (3), 420–424. doi: 10.1016/s0003-2697(69)80009-6

Verma, P., Goyal, R., Chahota, R., Sharma, T., Abdin, M., and Bhatia, S. (2015). Construction of a genetic linkage map and identification of QTLs for seed weight and seed size traits in lentil (*Lens culinaris* medik.). *PloS One* 10 (10), e0139666. doi: 10.1371/journal.pone.0139666

Verma, P., Sharma, T., Srivastava, P., Abdin, M., and Bhatia, S. (2014). Exploring genetic variability within lentil (*Lens culinaris* medik.) and across related legumes using a newly developed set of microsatellite markers. *Mol. Biol. Rep.* 41 (9), 5607–5625. doi: 10.1007/s11033-014-3431-z

Wang, N. (2008). Effect of variety and crude protein content on dehulling quality and on the resulting chemical composition of red lentil (*Lens culinaris*). *J. Sci. Food Ag.* 88, 885–890. doi: 10.1002/jsfa.3165

Wang, B., Jin, S. H., Hu, H. Q., Sun, Y. G., Wang, Y. W., Han, P., et al. (2012). UGT87A2, an *Arabidopsis* glycosyltransferase, regulates flowering time *via* FLOWERING LOCUS c. *New Phytol.* 194, 666–675. doi: 10.1111/j.1469-8137.2012.04107.x

Wang, D., Song, W., Wei, S., Zheng, Y., Chen, Z., Han, J., et al. (2018). Characterization of the ubiquitin c-terminal hydrolase and ubiquitin-specific protease families in rice (*Oryza sativa*). *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.01636

Xu, Y., Li, H. N., Li, G. J., Wang, X., Cheng, L. G., and Zhang, Y. M. (2011). Mapping quantitative trait loci for seed size traits in soybean (*Glycine max* l. *Merr.). Theor. Appl. Genet.* 122, 581–559. doi: 10.1007/s00122-010-1471-x

Yang, P., Smalle, J., Lee, S., Yan, N., Emborg, T. J., and Vierstra, R. D. (2007). Ubiquitin c-terminal hydrolases 1 and 2 affect shoot architecture in arabidopsis. *Plant J.* 51 (3), 441–457. doi: 10.1111/j.1365-313X.2007.03154.x

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 16 February 2023
DOI 10.3389/fpls.2023.1097001

OPEN ACCESS

EDITED BY
Zhengjun Xia,
Chinese Academy of Sciences (CAS), China

REVIEWED BY
Rongxia Guan,
Chinese Academy of Agricultural Sciences (CAAS), China
Dayong Zhang,
Nanjing Agricultural University, China

*CORRESPONDENCE
Yuxian Zhang
13836962211@126.com

SPECIALTY SECTION
This article was submitted to Functional and Applied Plant Genomics, a section of the journal Frontiers in Plant Science

RECEIVED 13 November 2022
ACCEPTED 02 February 2023
PUBLISHED 16 February 2023

CITATION
Zhang X, Zhong J, Cao L, Ren C, Yu G, Gu Y, Ruan J, Zhao S, Wang L, Ru H, Cheng L, Wang Q and Zhang Y (2023) Genome-wide characterization of aldehyde dehydrogenase gene family members in groundnut (*Arachis hypogaea*) and the analysis under saline-alkali stress. *Front. Plant Sci.* 14:1097001. doi: 10.3389/fpls.2023.1097001

Genome-wide characterization of aldehyde dehydrogenase gene family members in groundnut (*Arachis hypogaea*) and the analysis under saline-alkali stress

Xiaoming Zhang[1,2,3], Jingwen Zhong[1], Liang Cao[1,3], Chunyuan Ren[1,3], Gaobo Yu[1,3], Yanhua Gu[1], Jingwen Ruan[2], Siqi Zhao[2], Lei Wang[1], Haishun Ru[1], Lili Cheng[4], Qi Wang[5] and Yuxian Zhang[1,3]*

[1]Heilongjiang Bayi Agricultural University, Key Laboratory of Soybean Mechanized Production, Ministry of Agriculture and Rural Affairs, Daqing, China, [2]Agricultural College, Northeast Agricultural University, Harbin, China, [3]National Coarse Cereals Engineering Research Center, Heilongjiang Bayi Agricultural University, Daqing, China, [4]Institute of Industrial Crops, Heilongjiang Academy of Agricultural Sciences, Harbin, China, [5]Institute of Crop Cultivation and Tillage, Heilongjiang Academy of Agricultural Sciences, Harbin, China

Groundnut or peanut (*Arachis hypogaea*) is a legume crop. Its seeds are rich in protein and oil. Aldehyde dehydrogenase (ALDH, EC: 1.2.1.3) is an important enzyme involved in detoxification of aldehyde and cellular reactive oxygen species, as well as in attenuation of lipid peroxidation-meditated cellular toxicity under stress conditions. However, few studies have been identified and analyzed about ALDH members in *Arachis hypogaea*. In the present study, 71 members of the ALDH superfamily (AhALDH) were identified using the reference genome obtained from the Phytozome database. A systematic analysis of the evolutionary relationship, motif, gene structure, *cis*-acting elements, collinearity, Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment, and expression patterns was conducted to understand the structure and function of *AhALDH*s. *AhALDH*s exhibited tissue-specific expression, and quantitative real-time PCR identified significant differences in the expression levels of *AhALDH* members under saline-alkali stress. The results revealed that some *AhALDHs* members could be involved in response to abiotic stress. Our findings on *AhALDHs* provide insights for further study.

KEYWORDS

aldehyde dehydrogenase, evolutionary, *cis*-acting elements, expression pattern, saline-alkali stress

1 Introduction

Aldehyde molecules are essential intermediate compounds generated in catabolic and biosynthetic pathways during biological development and growth (Vasiliou et al., 2000). In response to stress, aldehyde accumulates in cells, causing an imbalance and interfering with cellular homeostatic metabolic responses (Bartels, 2001), and becomes toxic if present in

excess (Carmona-Molero et al., 2021) Aldehyde dehydrogenase (ALDH, EC: 1.2.1.3) as a scavenger of aldehyde molecules contributes to their homeostasis (Yoshida et al., 1998). The ALDH family is composed of a variety of $NAD(P)^+$-dependent enzymes that irreversibly oxidize endogenously and exogenously derived aldehyde molecules to carboxylic acids (Yoshiba et al., 1997). ALDH enzymes also function in intermediary metabolism by providing protection from osmotic stress and generating NAD(P)H (Kelly and Gibbs, 1973; Ishitani et al., 1995; Brocker et al., 2010). ALDHs have been reported to improve stress resistance in crops (Bartels and Sunkar, 2005). ALDH family members are found in prokaryotic and eukaryotic organisms and are highly conserved and well represented in virtually all plant species (Brocker et al., 2013). Research on the ALDH gene family in prokaryotes and mammals is abundant (Kirch et al., 2004). The ALDH members associated with ALDH enzyme activity are linked to diseases such as cataracts, hyperprolinaemia, and cancers (Jackson et al., 2011). However, the functional and structural characterization of plant ALDHs and gene duplication events underlying their diversification have lagged behind that of their mammalian and bacterial counterparts (Zhang et al., 2012).

ALDH family members, which are found in almost all plant species, are variable, widespread in plant tissues, and regulated developmentally (Chen C, et al., 2014; Guo et al., 2017). ALDH members participate in plant growth and development and play a vital role in catabolic and bio-synthetic pathways, such as carnitine biosynthesis (Marchitti et al., 2008), glycolysis/gluconeogenesis (Tylichová et al., 2010), and amino-acid metabolism (Yang et al., 2011). The *ALDH2B2* (*rf2*) gene has ALDH domain functions and is a male fertility restorer in maize (Skibbe et al., 2002). *ALDH7s* in *Arabidopsis* and soybean are involved in aldehyde detoxification (Shin et al., 2009), whereas *OsALDH7* is essential for seed maturation, and its mutation leads to seed browning during seed drying and storage of rice (Shin et al., 2009). An increasing number of studies have shown that some ALDH members indirectly function in plant cell protection under various abiotic stresses through detoxification of cellular reactive oxygen species (ROS) and/or reduction of lipid peroxidation (Shin et al., 2009; Singh et al., 2013). Betaine aldehyde dehydrogenases (BADH) are a type of ALDH10 enzyme that catalyze the oxidation of betaine aldehyde into glycine betaine (a major cellular osmolyte) and thereby improve plant resistance to environmental stress (Le Rudulier et al., 1984); the *BADH* gene has been shown to improve salt tolerance in plants (Zhou et al., 2008). Furthermore, the ectopic expression of *ALDH3I1* and *ALDH7B4* significantly reduced malondialdehyde (MDA) levels and lipid peroxidation in transgenic *Arabidopsis*, revealing the role of these two genes in increasing plant tolerance to drought and salt stress (Kotchoni et al., 2006). *VvALDH2B4*, an ALDH member from the Chinese wild grapevine (*Vitis pseudoreticulata*), lowered MDA levels and enhanced plant resistance to salt stress and pathogenic bacteria in over-expressed transgenic *Arabidopsis* (Wen et al., 2012). *ScALDH21*, an ALDH member isolated from *Syntrichia caninervis*, enhanced the activity of antioxidant enzymes, increased the proline content, and lowered the MDA content in transgenic tobacco under salt and drought stress, which is likely the reason for increased germination ratios and root lengths in tobacco plants (Yang et al., 2015).

ALDH members have been identified in *Arabidopsis* (Kirch et al., 2001), rice (*Oryza sativa*) (Gao and Han, 2009), maize (*Zea mays*) (Jimenez-Lopez et al., 2010), soybean (*Glycine max*) (Kotchoni et al., 2012), and cotton (*Gossypium hirsutum*) (Guo et al., 2017), but few studies have examined their presence in groundnut (*Arachis hypogaea*). Groundnut (peanut), a member of the legume family, originated in southern Bolivia (Krapovickas and Gregory, 1994). It is cultivated in more than 100 countries on 26 million hectares (ha) of land for its seeds that are a rich source of dietary fiber, minerals, vitamins, and bioactive compounds, especially proteins and oil (Desmae et al., 2019). The growth of groundnut is affected by abiotic stresses, and research has been conducted to increase its stress resistance through molecular breeding (Desmae et al., 2019). In the present study, ALDH members in groundnut were identified, and a comprehensive analysis (location, evolution, motif, gene structure, *cis*-acting elements, and expression patterns) was conducted. The results can be utilized to breed groundnut with improved stress resistance.

2 Materials and methods

2.1 ALDH members of *Arachis hypogaea*

The reference genome (*Arachis hypogaea* v1.0) and adjoining information (protein sequence, mRNA, coding sequence, and DNA) were obtained from the Phytozome database (Goodstein et al., 2012). The ALDH domain (accession number PF00171) was downloaded from the PFAM database. The ALDH members were identified through hmmsearch and hmmbuild using the perl script in the Linux system, in which $1e^{-15}$ was set as a filter threshold (Zhang et al., 2020). The SMART software (Letunic and Bork, 2018) was used to confirm the ALDH domain and remove duplicates.

2.2 Analysis of ALDH members

The evolutionary relationship of ALDH members was analyzed using neighbor-joining methods and a Poisson model implemented in MEGA X, with 1000 bootstrap repetitions (Kumar et al., 2018). The motif of ALDH members was analyzed using the MEME platform, in which the length of motifs was 10–15 amino acids, while the e-value of motifs was less than e^{-5} (Bailey et al., 2009). The gene structure of the ALDH members was analyzed using the GSDS software (Hu et al., 2015). The *cis*-acting elements were identified and predicted using PlantCare, in which the function of each *cis*-acting element was predicted (Lescot et al., 2002). The expression of ALDH members and their location on chromosomes was extracted from the Phytozome database (Goodstein et al., 2012). The results were plotted using TBtool (Chen et al., 2020). The Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases were used for the gene annotation of ALDH members (Chen et al., 2017); Majorbio Cloud provided the platform for enrichment analysis (Shanghai, China).

2.3 Plant materials and conditions

The locally grown groundnut cultivar Silihong was used in this study. The seeds were provided by the Institute of Economic Botany, Heilongjiang Academy of Agricultural Sciences (Harbin, Hei Longjiang, China). Seeds of the same size were sterilized with 5% NaClO solution for 15 min, rinsed with distilled water, and placed in a Petri dish lined with double-layer filter paper. The seeds were covered with a single filter paper. Three replicates of 10 seeds per Petri dish were prepared for saline-alkali and control treatments. Petri dishes were placed in a GXZ intelligent light incubator (Ningbo Jiangnan Instrument Factory, Zehjiang, China) at 25°C in the dark. When the length of the peanut buds reached half the length of the peanut seeds, the seeds were sampled and used for expression pattern analysis. Control seeds were treated with distilled water. The rest of the seeds were subjected to saline-alkali treatment by adding saline-alkali solution (pH, 8.9) to Petri dishes. The sprouts in each Petri dish were set as the experimental unit, while each treatment had three experimental units as biological replicates.

2.4 Physiological and expression level analysis

The radicle was used as samples in this study, with biological and technological replicates. ALDH enzyme activity was determined under water and saline-alkali treatment at 0, 12, 24, 48, 72, and 96 h using a Multiskan FC microplate reader (Thermo Scientific Company, Wilmington, DE, USA) and the Elisa Kit (M0608, Michy Biology, Suzhou, China) according to the manufacturer's instructions. Radicles were used as samples. The samples at 0 and 48 h from both treatments were used for quantitative real-time PCR (qRT-PCR) analysis. Total RNA was extracted using a MolPure® Plant Plus RNA Kit (19292ES, Yeason, Shanghai, China). The qualified RNA was used for reverse transcription using a kit (11139ES, Yeason) after detection using a Nanodrop OneC (Thermo Fisher Scientific, Waltham, MA, USA) and 1% agarose for RNA quality. The primers for ALDH amplification were designed using DNAMAN (Lynnon Biosoft, San Ramon, CA, USA) (Nong et al., 2019), and the *UKN1* gene was used as the reference gene (Table S2). qRT-PCR was performed on the Roche platform (480 II, Roche, Basel, Switzerland) using Hieff UNICON® Universal Blue qPCR SYBR Green Master Mix (11184ES, Yeason) according to the manufacturer's instructions. The expression levels of ALDH members were calculated using the $2^{-\Delta Ct}$ method (Livak and Schmittgen, 2001). Data were analyzed in the SPSS software (Bezzaouha et al., 2020). Duncan's multiple method was used to test differences between groups.

3 Results

3.1 Identification and evolution of ALDH members

One hundred and five ALDH members were identified from the reference genome (*Arachis hypogaea* v1.0) from the Phytozome database. The sequences were confirmed using the Smart software (Letunic and Bork, 2018), and after removing the duplicates, 71 members of the ALDH family were identified. The evolutionary relationships of the AhALDH members were analyzed using the MEGA X software (Kumar et al., 2018), while ALDH members in *Arachis duranensis* and *Arachis ipaensis* were used to correct evolutionary relationships. All these ALDH members were classified into 10 subfamilies, labelled I–X according to the order of AhALDH evolution. Subfamilies III and VI had the fewest AhALDH members, only two, while subfamily X had the highest number of AhALDH members at 17. These AhALDH members were named according to the evolutionary results and Brockers' method (Brocker et al., 2013). ALDH members in *A. duranensis* and *A. ipaensis*, along with AhALDH members, were also divided into 10 subfamilies, which proved the accuracy of the evolutionary analysis (Figure 1). Moreover, detailed information about the proteins (including protein length, molecular weight, and isoelectric point) is shown in Table S1. The protein length varied from 303 to 791 amino acids, the molecular weight was 32.48–88.83 kDa, and the isoelectric point was 4.91–9.61. This information revealed that AhALDH members in different subgroups had diversified features, indicating that different members might perform different functions.

3.2 Location analysis of AhALDHs

Location information was obtained from databases and drawn using Tbtools (Figure 2). All chromosomes had AhALDH members, except for Chr 10 and Chr 20. The number of AhALDH members on each Chr differed: Chr13 and Chr15 had the most AhALDH members (9), while Chr03 and Chr11 had seven members; Chr1, Chr7, and Chr12, with only one member each, had the fewest, while only two AhALDH members were found on Chr16 and Chr19. These results indicated that the distribution of AhALDH members on groundnut chromosomes is not uniform, indicating that Chr13 and Chr15 might be central to AhALDH evolution.

3.3 Motif and gene structure analysis of AhALDHs

The motif and gene structure of AhALDHs were analyzed using MEME and GSDS, respectively, to explore their structure and properties. The results are shown in Figure 3. The length of the motifs varied from 10 to 50, and the e-value was lower than $1e^{-20}$. AhALDHs in the same subfamily shared similar kinds of motifs, although some motifs also varied within the same subfamily, such as motif 1 (Figure 3). Members of the same subfamily had a similar gene structure, with the longest AhALDH member being in subfamily X. Moreover, the motif and gene structure of ALDH members in *A. duranensis* and *A. ipaensis* also revealed that members in each subfamily had similar characteristics, indicating that they might perform similar functions.

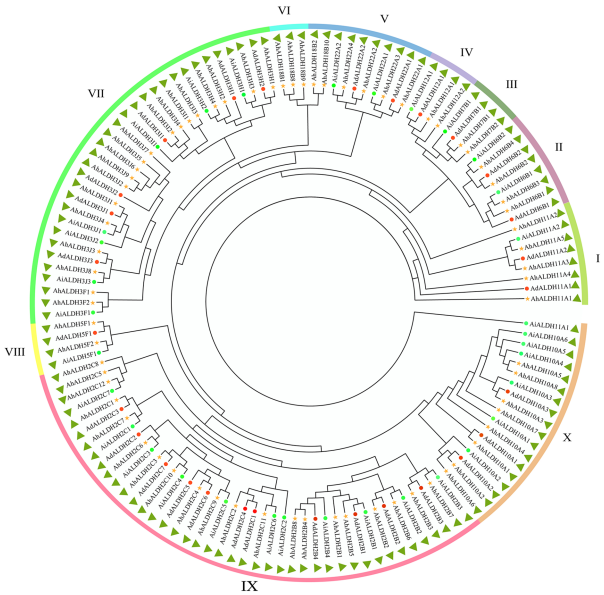


FIGURE 1
Evolutionary relationship of AhALDH members. Green circles represent ALDH members in *Arachis hypogaea*, while red circles and yellow stars represent ALDH members in *Arachis duranensis* and *Arachis ipaensis*. Outer ring bands in different colors represent the subfamilies I–X.

3.4 Comparison of ALDHs in different species

In this study, 260 ALDH members in seven species, namely groundnut, *Arabidopsis*, rice, human, cotton, soybean, and grape (*Vitis vinifera*), were used for the evolutionary and motif analysis. These 260 members were divided into 10 subfamilies, in which the members within each subfamily shared similar motifs. The evolutionary results were similar to those shown in Figure 2. Under the same analysis threshold, the AhALDHs had similar kinds of motifs to ALDH members in other species (Figure 4).

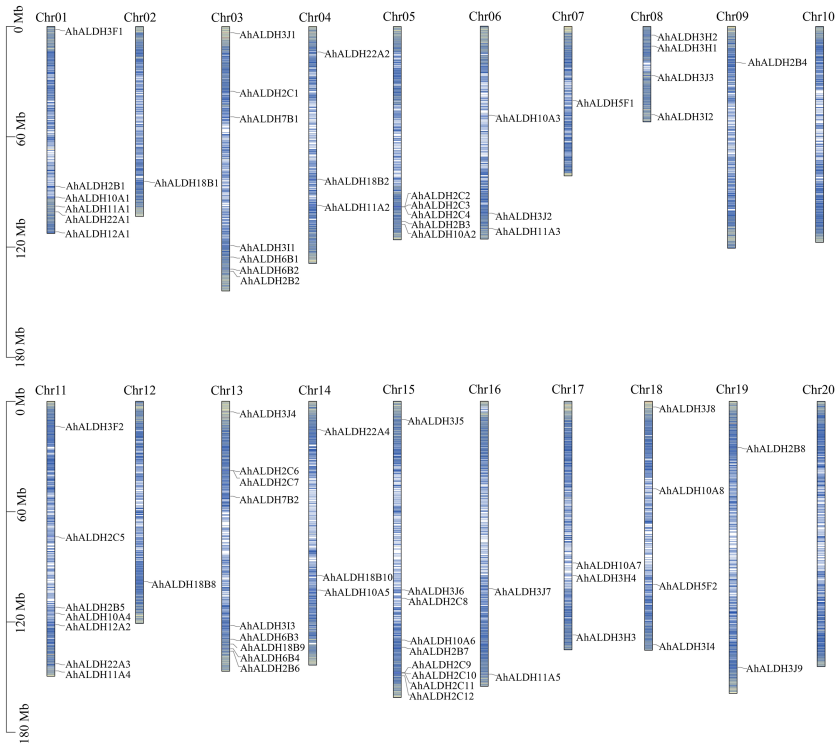


FIGURE 2
Location of *AhALDH* members on chromosomes (Chr1–Chr20). The length of the columns represents the size of the chromosomes. The black scale on the left represents the position, and blue lines inside the columns represent the gene density on chromosomes.

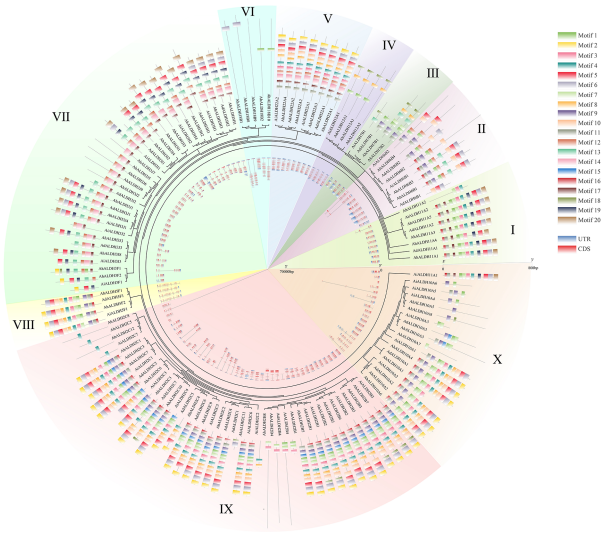


FIGURE 3
Motif and gene structure analysis of AhALDH members. The middle ring represents the evolutionary relationship of AhALDHs. The 10 families are highlighted with differently colored backgrounds. The outer ring shows the motif analysis of AhALDHs. Motifs from 1 to 20 are marked with differently colored boxes. The inner ring displays the gene structure analysis of AhALDHs. Blue boxes are untranslated regions (UTRs) and pink boxes are coding sequences (CDSs), black lines are the intron region.

3.5 Cis-acting element analysis of AhALDHs

The *cis*-acting elements of *AhALDH*s were analyzed and their function predicted using the Plantcare software (Figure 5; Table S3). Three kinds of elements were recognized based on their function. The *cis*-acting elements TGA-element, ABRE, P-box, GARE-motif, TCA-element, AT-rich sequence, SARE, TATC-box, and AuxRR-core (marked in red in Figure 4) were predicted to be related to hormones; LTR, MBS, ARE, and GC-motif (marked in blue) were related to abiotic stress. The members within each subfamily had similar *cis*-acting elements, and the analysis revealed that *AhALDH*s might be associated with hormones and abiotic stress; in particular, some members had stress-related *cis*-acting elements.

3.6 Collinearity analysis of AhALDHs

Fifty-two pairs of collinear *AhALDH*s were identified. Most of the *AhALDH* members had more than two collinear pairs; *AhALDH2B3* had five collinear pairs, which revealed that *AhALDH2B3* might have a longer and more pivotal evolutionary relationship (Figure 6A). Compared with the *Arabidopsis* members, *AhALDH*s had 24 collinear pairs and likely similar functions to those of *Arabidopsis* collinear

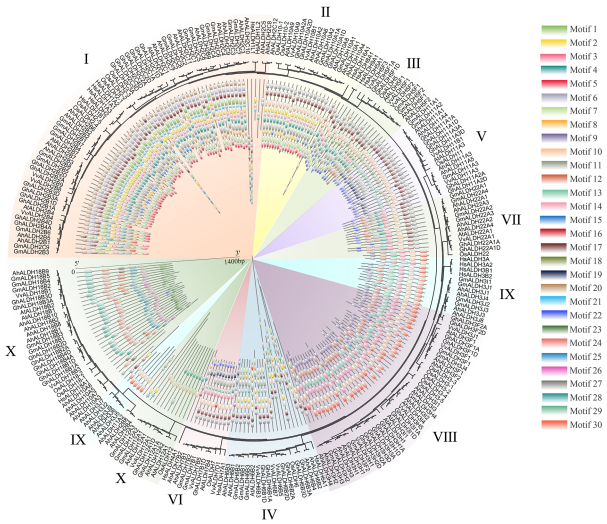


FIGURE 4
Analysis of ALDH members in seven species. The seven species were groundnut (*Ah*), *Arabidopsis* (*At*), rice (*Os*), human (*Hs*), cotton (*Gh*), soybean (*Gm*), and grape (*Vv*). Differently colored backgrounds represent different subfamilies. The inner circle represents the motifs of ALDH members; different squares show different motifs.

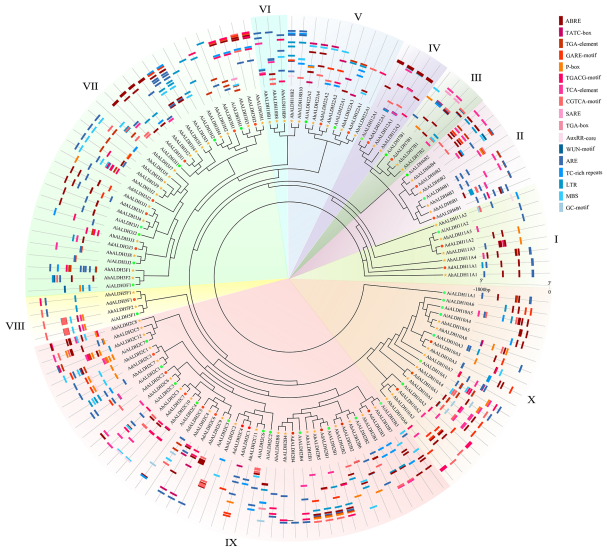


FIGURE 5
Cis-acting element analysis of *AhALDH* members. The inner ring shows the evolutionary relationship of AhALDHs; differently colored backgrounds represent different subfamilies. The outer ring shows the *Cis*-acting elements analysis. Red, orange, and pink boxes in different shapes represent *cis*-acting elements associated with hormones, while the blue boxes are *cis*-acting elements responding to abiotic stress.

genes. Interestingly, both collinear *Arabidopsis* members and *AhALDH* members were not distributed on the last chromosome, which potentially reveals some characteristics of the ALDH members (Figures 2, 6B).

3.7 GO and KEGG enrichment analysis of AhALDHs

The GO and KEGG enrichment analyses (Figure 7) were conducted to understand the function of *AhALDH*s. The *AhALDH*s were significantly ($P < 0.05$) enriched in 81 GO terms (Table S4); the top-10 GO terms were enriched in semialdehyde dehydrogenase activity (6) and amino acid-related terms (4), suggesting that *AhALDH*s might participate in the growth of plants (Figure 7A). The analysis further revealed that the *AhALDH*s were significantly ($P < 0.05$) enriched in 20 pathways (Table S5); eight pathways were related to amino acid metabolism, indicating the role of *AhALDH*s in the biosynthesis of amino acids (Figure 7B). These results indicated that *AhALDH*s might exercise function through amino acid pathways.

3.8 Tissue-specific expression pattern analysis of *AhALDH*s

The expression pattern of 28 *AhALDH* members from different tissues encompassing all subfamilies was extracted from the Phytozome database, and a heatmap was drawn using TBtools. The

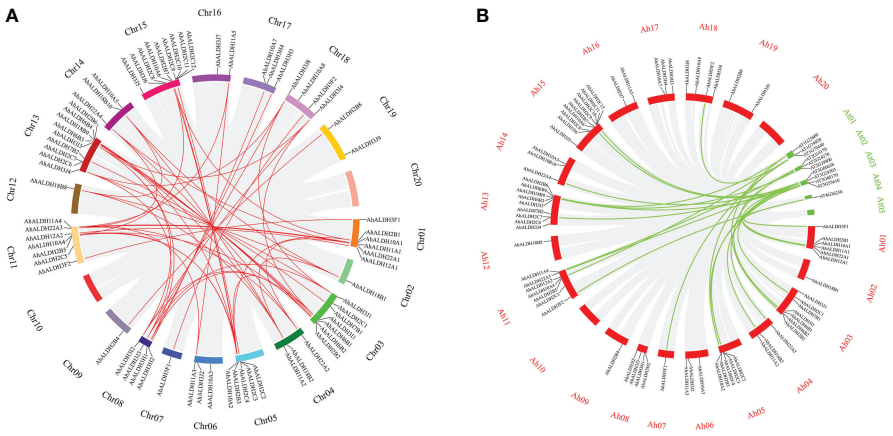


FIGURE 6
Collinear analysis of *AhALDH* members in groundnut and *Arabidopsis*. **(A)** Collinear analysis of *AhALDH* members in *Arachis hypogaea*. The circles of different colors represent different chromosomes, while red lines represent collinear pairs in *AhALDH*s, and a gray background represents all collinear blocks. **(B)** Collinear analysis of *AhALDH*s with *Arabidopsis* members. The red circles represent groundnut chromosomes, while green circles represent *Arabidopsis* chromosomes, and the green lines represent collinear gene pairs.

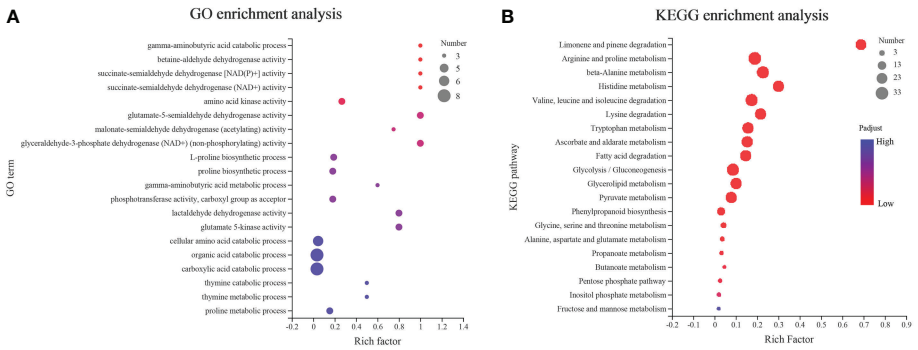


FIGURE 7
GO and KEGG enrichment analysis of *AhALDH* members. The size of each circle represents the number of *AhALDH*s. The transition in color from blue to red represents the P-value from high to low. **(A)** GO enrichment analysis of *AhALDH*s. **(B)** KEGG enrichment analysis of *AhALDH*s.

results showed that the expression of *AhALDHs* was tissue-specific (Figure 8). Some *AhALDHs* had a higher expression level (such as *AhALDH12A1*, *AhALDH18B1*, *AhALDH6B2*, *AhALDH2C7*, *AhALDH6B4*, *AhALDH2C11*, and *AhALDH11A5*) in roots than in other tissues; therefore, roots should be used as a target tissue of *AhALDHs* for further study.

3.9 ALDH activity under saline-alkali stress

The ALDH activity was determined under control and stress conditions at 0, 12, 24, 48, 72, and 96 h (Figure 9). The ALDH activity showed no obvious trend under the control conditions, but the data of activity under saline-alkali stress had increased. The results showed that the ALDH activity began to significantly change at 48 h ($P < 0.05$). All these results illustrated that 48 h could be used as an simulated stress time for testing the expression of *AhALDH* members according to ALDH Activity.

3.10 Expression analysis under saline-alkali stress

Twelve *AhALDHs* with higher expression in the roots at the sprout stage, based on the tissue-specific expression pattern analysis, were selected for qRT-PCR analysis under stress, and 48 h was used as the sampling time. The expression of these members significantly differed between the control and stress treatment ($P < 0.05$). Some *AhALDH* members, *AhALDH10A1*, *AhALDH22A1*, *AhALDH12A1*, *AhALDH6B2*, *AhALDH3H2*, *AhALDH3H1*, *AhALDH10A4*, *AhALDH2C11*, *AhALDH11A5*, and *AhALDH3H3*, were upregulated under saline-alkali stress, while others, *AhALDH22A4* and *AhALDH3H4*, were downregulated. These results suggested the role of these members in plant response to abiotic stress; some might be involved in positive regulation, while some might be involved in negative regulation (Figures 10A-L).

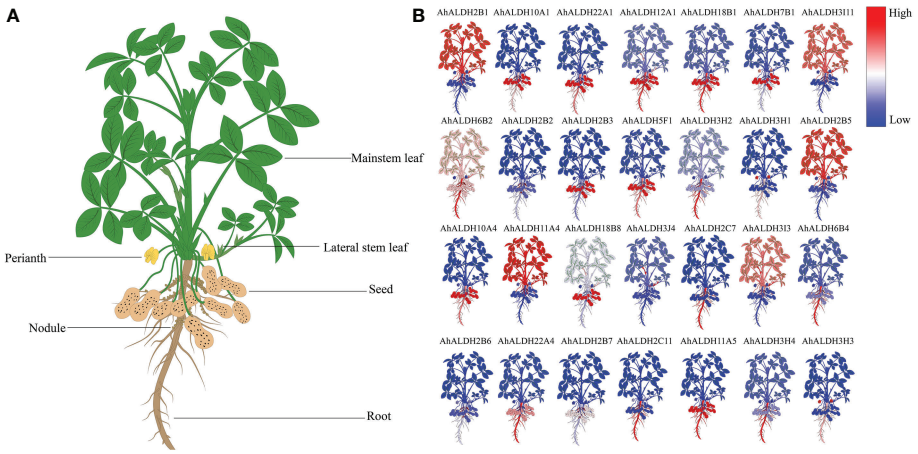


FIGURE 8
Expression pattern analysis of *AhALDH*s in different tissues. **(A)** Schematic diagram of different groundnut tissues; plants were divided into root, nodule, seed, perianth, and leaf. **(B)** Expression levels of *AhALDH*s. A change in color from blue to red represents *ALDH* levels from low to high.

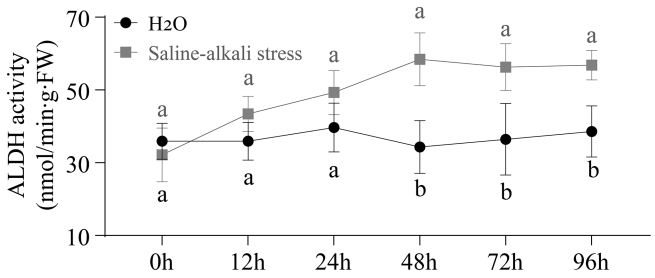


FIGURE 9
Variation curve of ALDH enzyme activity under control and saline-alkali stress treatment. Black line represents ALDH enzyme activity in H_2O treatment at the sprout stage, while gray line represents ALDH enzyme activity in saline-alkali stress treatment.

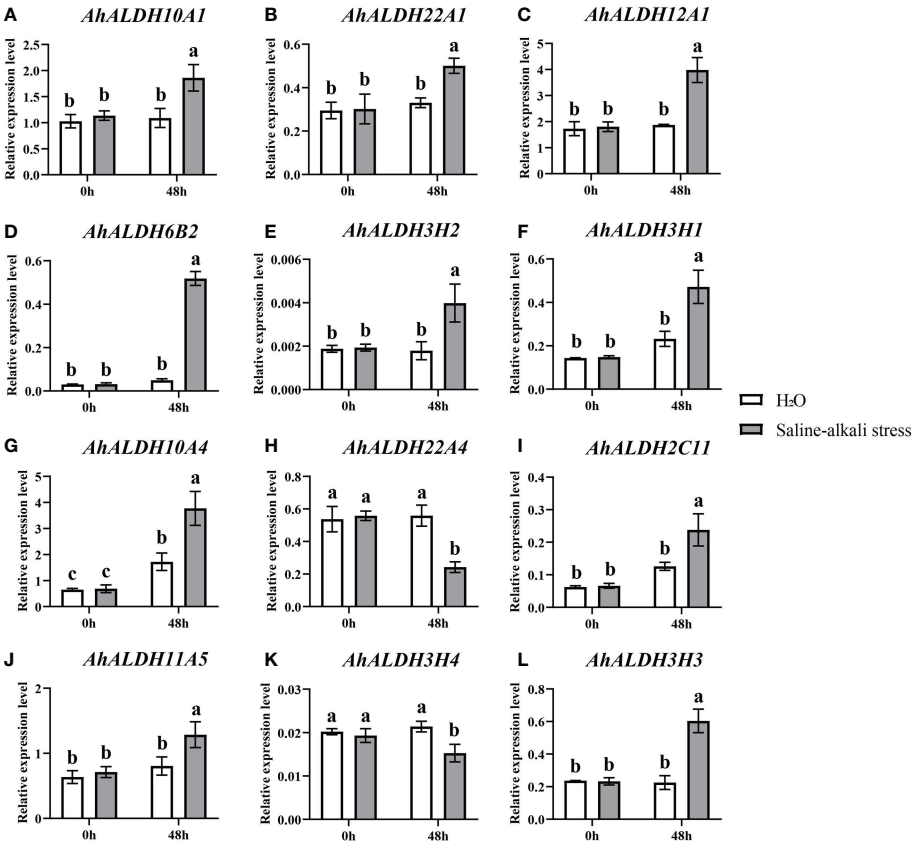


FIGURE 10
Expression pattern analysis of 12 *AhALDHs* under saline-alkali stress at 0h and 48h. White columns show the expression in the H_2O at the sprout stage, while gray columns show the expression in the saline-alkali stress treatment. **(A–L)** The expression analysis of 12 AhALDH members.

4 Discussion

Members of the ALDH superfamily have previously been identified in several species (Brocker et al., 2013): 9 ALDH members were identified in *Arabidopsis* (Kirch et al., 2004), 20 were reported for rice (Gao and Han, 2009), 23 for grape (Zhou et al., 2012), and 19 for tomato (*Solanum lycopersicum*) (Zhang et al., 2012); 28 for maize (Zhou et al., 2012), 19 for sorghum (*Sorghum bicolor*) (Islam et al., 2022), and 53 for soybean (Wang et al., 2017). In the present study, 71 ALDH members were identified through the reference genome (*Arachis hypogaea* v1.0). Such a large number of ALDH members might be related to the large size of the genome and the evolution of the genes (Wang et al., 2017). The evolutionary relationship analysis resolved these members into 10 subfamilies (Man et al., 2020); similarly, *VvALDHs*, *ZmALDHs*, and *GhALDHs* were divided into 10 subfamilies (Zhang et al., 2012; Zhou et al., 2012; Dong et al., 2017), which reveals that 10 subfamilies express the evolutionary relationship characteristics of ALDH members. The evolutionary relationship of seven species confirmed the accuracy of the member classification into 10 subfamilies.

Motif analysis can be used to reveal the function of members according to protein sequences, in which the similar kinds of motifs may have a similar function (Bailey et al., 2015; Liu et al., 2021). The motif analysis of AhALDHs showed that the members in each subfamily shared similar kinds of motifs; these results are in accordance with those reported for ALDH members in soybean (Wang et al., 2017). In *cis*-acting element analysis, *AhALDH*s with TGA-element, ABRE, P-box, GARE-motif, TCA-element, AT-rich sequence, SARE, TATC-box, and AuxRR-core were related to hormones, whereas LTR, MBS, ARE, and GC-motif elements were associated with abiotic stress response, suggesting that *AhALDH*s are related to abiotic stress and hormones. Similarly, the elements LTR, MBS, ABRE, P-box, TCA-element, and AuxRR-core were found in *SbALDHs* (Islam et al., 2022), whereas LTR, MBS, ARE, ARBE, and TCA-element were predicted in *GhALDH*s (Guo et al., 2017; Yang et al., 2019). Thus, *cis*-acting elements such as LTR, MBS, ARE, ARBE, and TCA-element are common *cis*-acting elements among ALDH members, contributing to the role of *AhALDH*s in plant response to abiotic stress and hormone interactions (Chen Z, et al., 2014). Some *AhALDH*s were colinear with *Arabidopsis* genes; for example, *AhALDH10A2* was colinear with *AT1G74920*, which is involved in plant response to salt and drought (Missihoun et al., 2011); *AhALDH18B9* was colinear with *AT2G39800* (*P5CS1*), which acts in abiotic stress response (salt, oxidative, ABA, desiccation, and water deprivation) and participates in proline biosynthesis (Yoshiba et al., 1995; Yoshiba et al., 1999; Huang et al., 2008; Székely et al., 2008); and *AhALDH2B2* and *AhALDH2B6* were colinear with *AT1G23800* (*ALDH2B7*), a gene that responds to drought and abscisic acid signal (Depuydt and Vandepoele, 2021). These relationships indicated the role *AhALDH*s play in plant response to abiotic stresses. *AhALDH*s were enriched in semialdehyde dehydrogenase activity (GO results) and amino acid metabolism (GO and KEGG analysis). Semialdehyde dehydrogenase activity is closely related with abiotic stress and increases stress conditions (Wei et al., 2021; Zhang et al., 2021). Amino acid metabolism usually intensifies during the growth and development of plants especially under stress, such as GST family members (Batista-Silva et al., 2019; Yang et al., 2019). Some amino acids, such as proline, have been used as a parameter for assessing damage to plants under abiotic stress (Ghosh et al., 2022). The enrichment analysis further corroborated the correlation between *AhALDHs* and abiotic stress.

The expression of several ALDH members, such as *GhALDH*s, *GmALDH*s, and *SiALDH*s, shows a tissue-specific pattern and is higher in roots (Dong et al., 2017; Wang et al., 2017). A similar pattern was observed in the present study, suggesting that the root system is a suitable target for studying ALDH members. Some ALDH members, such as *BrALDH7B2*, were predicted to function in stress response (Gautam et al., 2019). Similar to that of *VvALDHs*, *SiALDHs*, and *CaALDHs* (Zhang et al., 2012; Chen C, et al., 2014; Carmona-Molero et al., 2021), the expression of some *AhALDHs* (such as *AhALDH10A1*, *AhALDH22A1*, *AhALDH12A1*, or *AhALDH6B2*) was significantly upregulated, while it was downregulated in others (such as *AhALDH22A4* and *AhALDH3H4*) under alkali-stress conditions, which these upregulated six members could be considered as cluster genes involved in resistance to abiotic stress.

5 Conclusion

In this study, 71 ALDH members were identified from the groundnut reference genome and classified into 10 subfamilies, with similar motifs and gene structure. *AhALDHs* were associated with abiotic stress response and hormones *via cis*-acting elements. The results of collinearity and enrichment analysis (including GO and KEGG pathway analysis) revealed that *AhALDHs* are involved in plant response to stress and their expression is tissue-specific. The root system is a target tissue suitable for studying *AhALDHs*, and upregulated members can be used as candidate *AhALDHs* members (such *AhALDH10A1*, *AhALDH22A1*, *AhALDH12A1*, or *AhALDH6B2*) involved in resistance to abiotic stress in future research. This study provides insights into *AhALDHs* and the basis for further research of groundnut.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

Author contributions

Conceptualization: XZ and JZ; methodology: LCa and GY; software: CR; validation: YG, JR, HR, and SZ; formal analysis and investigation: LCh and YZ; writing—review and editing: LW, XZ and YZ; funding acquisition: YZ. All authors contributed to the article and approved the submitted version.

Funding

This research was funded by Heilongjiang Province's "Revealing the List and Commanding the Leaders" scientific and technological research project (2021ZXJ05B02); China Agriculture Research System of MOF and MARA (CARS-04-PS18); Heilongjiang Bayi Agricultural University Support Program for San Heng San Zong (TDJH202001); Natural Science Foundation of Heilongjiang Province (LH2022115).

Acknowledgments

We thank Heilongjiang Academy of Agricultural Sciences for providing plant materials and Northeast Agricultural University for providing instrument platforms.

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1097001/full#supplementary-material

SUPPLEMENTARY TABLE 1
The primers sequence of ALDH members in groundnut.

SUPPLEMENTARY TABLE 2
The detailed information of ALDH members in groundnut.

SUPPLEMENTARY TABLE 3
The information of *cis*-acting elements.

SUPPLEMENTARY TABLE 4
The Gene Ontology (GO) terms enriched in *AhALDHs*.

SUPPLEMENTARY TABLE 5
The Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways enriched in *AhALDHs*.

References

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. doi: 10.1093/nar/gkp335

Bailey, T. L., Johnson, J., Grant, C. E., and Noble, W. S. (2015). The MEME suite. *Nucleic Acids Res.* 43, W39–W49. doi: 10.1093/nar/gkv416

Bartels, D. (2001). Targeting detoxification pathways: an efficient approach to obtain plants with multiple stress tolerance? *Trends Plant Sci.* 6, 284–286. doi: 10.1016/s1360-1385(01)01983-5

Bartels, D., and Sunkar, R. (2005). Drought and salt tolerance in plants. *Crit. Rev. Plant Sci.* 24, 23–58. doi: 10.1080/07352680590910410

Batista-Silva, W., Heinemann, B., Rugen, N., Nunes-Nesi, A., Araújo, W. L., Braun, H. P., et al. (2019). The role of amino acid metabolism during abiotic stress release. *Plant Cell Environ.* 42, 1630–1644. doi: 10.1111/pce.13518

Bezzaouha, A., Bouamra, A., Ammimer, A., and Ben Abdelaziz, A. (2020). Non-parametric tests on SPSS to compare two or more means on matched samples. *Tunis. Med.* 98 (12), 932–941.

Brocker, C., Lassen, N., Estey, T., Pappa, A., Cantore, M., Orlova, V. V., et al. (2010). Aldehyde dehydrogenase 7A1 (ALDH7A1) is a novel enzyme involved in cellular defense against hyperosmotic stress. *J. Biol. Chem.* 285, 18452–18463. doi: 10.1074/jbc.M109.077925

Brocker, C., Vasiliou, M., Carpenter, S., Carpenter, C., Zhang, Y., Wang, X., et al. (2013). Aldehyde dehydrogenase (ALDH) superfamily in plants: gene nomenclature and comparative genomics. *Planta* 237, 189–210. doi: 10.1007/s00425-012-1749-0

Carmona-Molero, R., Jimenez-Lopez, J. C., Caballo, C., Gil, J., Millán, T., and Die, J. V. (2021). Aldehyde dehydrogenase 3 is an expanded gene family with potential adaptive roles in chickpea. *Plants (Basel)* 10 (11), 2429. doi: 10.3390/plants10112429

Chen, Z., Chen, M., Xu, Z. S., Li, L. C., Chen, X. P., and Ma, Y. Z. (2014). Characteristics and expression patterns of the aldehyde dehydrogenase (ALDH) gene superfamily of foxtail millet (Setaria italica l.). *PloS One* 9, e101136. doi: 10.1371/journal.pone.0101136

Chen, C., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y., et al. (2020). TBtools: An integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant* 13, 1194–1202. doi: 10.1016/j.molp.2020.06.009

Chen, C. H., Ferreira, J. C., Gross, E. R., and Mochly-Rosen, D. (2014). Targeting aldehyde dehydrogenase 2: new therapeutic opportunities. *Physiol. Rev.* 94, 1–34. doi: 10.1152/physrev.00017.2013

Chen, L., Zhang, Y. H., Wang, S., Zhang, Y., Huang, T., and Cai, Y. D. (2017). Prediction and analysis of essential genes using the enrichments of gene ontology and KEGG pathways. *PloS One* 12, e0184129. doi: 10.1371/journal.pone.0184129

Depuydt, T., and Vandepoele, K. (2021). Multi-omics network-based functional annotation of unknown arabidopsis genes. *Plant J.* 108, 1193–1212. doi: 10.1111/tpj.15507

Desmae, H., Janila, P., Okori, P., Pandey, M. K., Motagi, B. N., Monyo, E., et al. (2019). Genetics, genomics and breeding of groundnut (Arachis hypogaea l.). *Plant Breed* 138, 425–444. doi: 10.1111/pbr.12645

Dong, Y., Liu, H., Zhang, Y., Hu, J., Feng, J., Li, C., et al. (2017). Comparative genomic study of ALDH gene superfamily in gossypium: A focus on gossypium hirsutum under salt stress. *PloS One* 12, e0176733. doi: 10.1371/journal.pone.0176733

Gao, C., and Han, B. (2009). Evolutionary and expression study of the aldehyde dehydrogenase (ALDH) gene superfamily in rice (Oryza sativa). *Gene* 431, 86–94. doi: 10.1016/j.gene.2008.11.010

Gautam, R., Ahmed, I., Shukla, P., Meena, R. K., and Kirti, P. B. (2019). Genome-wide characterization of ALDH superfamily in brassica rapa and enhancement of stress tolerance in heterologous hosts by BrALDH7B2 expression. *Sci. Rep.* 9, 7012. doi: 10.1038/s41598-019-43332-1

Ghosh, U. K., Islam, M. N., Siddiqui, M. N., Cao, X., and Khan, M. A. R. (2022). Proline, a multifaceted signalling molecule in plant responses to abiotic stress: understanding the physiological mechanisms. *Plant Biol. (Stuttg)* 24, 227–239. doi: 10.1111/plb.13363

Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., et al. (2012). Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* 40, D1178–D1186. doi: 10.1093/nar/gkr944

Guo, X., Wang, Y., Lu, H., Cai, X., Wang, X., Zhou, Z., et al. (2017). Genome-wide characterization and expression analysis of the aldehyde dehydrogenase (ALDH) gene superfamily under abiotic stresses in cotton. *Gene* 628, 230–245. doi: 10.1016/j.gene.2017.07.034

Hu, B., Jin, J., Guo, A. Y., Zhang, H., Luo, J., and Gao, G. (2015). GSDS 2. 0: an upgraded gene feature visualization server. *Bioinformatics* 31, 1296–1297. doi: 10.1093/bioinformatics/btu817

Huang, D., Wu, W., Abrams, S. R., and Cutler, A. J. (2008). The relationship of drought-related gene expression in arabidopsis thaliana to hormonal and environmental factors. *J. Exp. Bot.* 59, 2991–3007. doi: 10.1093/jxb/ern155

Ishitani, M., Nakamura, T., Han, S. Y., and Takabe, T. (1995). Expression of the betaine aldehyde dehydrogenase gene in barley in response to osmotic stress and abscisic acid. *Plant Mol. Biol.* 27, 307–315. doi: 10.1007/bf00020185

Islam, M. S., Mohtasim, M., Islam, T., and Ghosh, A. (2022). Aldehyde dehydrogenase superfamily in sorghum: genome-wide identification, evolution, and transcript profiling during development stages and stress conditions. *BMC Plant Biol.* 22, 316. doi: 10.1186/s12870-022-03708-4

Jackson, B., Brocker, C., Thompson, D. C., Black, W., Vasiliou, K., Nebert, D. W., et al. (2011). Update on the aldehyde dehydrogenase gene (ALDH) superfamily. *Hum. Genomics* 5, 283–303. doi: 10.1186/1479-7364-5-4-283

Jimenez-Lopez, J. C., Gachomo, E. W., Seufferheld, M. J., and Kotchoni, S. O. (2010). The maize ALDH protein superfamily: linking structural features to functional specificities. *BMC Struct. Biol.* 10, 43. doi: 10.1186/1472-6807-10-43

Kelly, G. J., and Gibbs, M. (1973). A mechanism for the indirect transfer of photosynthetically reduced nicotinamide adenine dinucleotide phosphate from chloroplasts to the cytoplasm. *Plant Physiol.* 52, 674–676. doi: 10.1104/pp.52.6.674

Kirch, H. H., Bartels, D., Wei, Y., Schnable, P. S., and Wood, A. J. (2004). The ALDH gene superfamily of arabidopsis. *Trends Plant Sci.* 9, 371–377. doi: 10.1016/j.tplants.2004.06.004

Kirch, H. H., Nair, A., and Bartels, D. (2001). Novel ABA- and dehydration-inducible aldehyde dehydrogenase genes isolated from the resurrection plant craterostigma plantagineum and arabidopsis thaliana. *Plant J.* 28, 555–567. doi: 10.1046/j.1365-313x.2001.01176.x

Kotchoni, S. O., Jimenez-Lopez, J. C., Kayodé, A. P., Gachomo, E. W., and Baba-Moussa, L. (2012). The soybean aldehyde dehydrogenase (ALDH) protein superfamily. *Gene* 495, 128–133. doi: 10.1016/j.gene.2011.12.035

Kotchoni, S. O., Kuhns, C., Ditzer, A., Kirch, H. H., and Bartels, D. (2006). Over-expression of different aldehyde dehydrogenase genes in arabidopsis thaliana confers tolerance to abiotic stress and protects plants against lipid peroxidation and oxidative stress. *Plant Cell Environ.* 29, 1033–1048. doi: 10.1111/j.1365-3040.2005.01458.x

Krapovickas, A., and Gregory, W. C. (1994). Taxonomıa del genero arachis (Leguminosae). *Bonplandia* 8, 1–186. doi: 10.30972/bon.160158

Kumar, S., Stecher, G., Li, M., Knyaz, C., and K. Tamura, (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096

Le Rudulier, D., Strom, A. R., Dandekar, A. M., Smith, L. T., and Valentine, R. C. (1984). Molecular biology of osmoregulation. *Science* 224, 1064–1068. doi: 10.1126/science.224.4653.1064

Lescot, M., Déhais, 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

Letunic, I., and Bork, P. (2018). 20 years of the SMART protein domain annotation resource. *Nucleic Acids Res.* 46, D493–d496. doi: 10.1093/nar/gkx922

Liu, Y., Zeng, Z., Zhang, Y. M., Li, Q., Jiang, X. M., Jiang, Z., et al. (2021). An angiosperm NLR atlas reveals that NLR gene reduction is associated with ecological specialization and signal transduction component deletion. *Mol. Plant* 14, 2015–2031. doi: 10.1016/j.molp.2021.08.001

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-delta delta C(T)) method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262

Man, J., Gallagher, J. P., and Bartlett, M. (2020). Structural evolution drives diversification of the large LRR-RLK gene family. *New Phytol.* 226, 1492–1505. doi: 10.1111/nph.16455

Marchitti, S. A., Brocker, C., Stagos, D., and Vasiliou, V. (2008). Non-P450 aldehyde oxidizing enzymes: the aldehyde dehydrogenase superfamily. *Expert Opin. Drug Metab. Toxicol.* 4, 697–720. doi: 10.1517/17425255.4.6.697

Missihoun, T. D., Schmitz, J., Klug, R., Kirch, H. H., and Bartels, D. (2011). Betaine aldehyde dehydrogenase genes from arabidopsis with different sub-cellular localization affect stress responses. *Planta* 233, 369–382. doi: 10.1007/s00425-010-1297-4

Nong, X., Zhong, S. N., Li, S. M., Yang, Y. J., Liang, Z., and Xie, Y. (2019). Genetic differentiation of pseudoregma bambucicola population based on mtDNA COII gene. *Saudi J. Biol. Sci.* 26, 1032–1036. doi: 10.1016/j.sjbs.2019.04.016

Shin, J. H., Kim, S. R., and An, G. (2009). Rice aldehyde dehydrogenase7 is needed for seed maturation and viability. *Plant Physiol.* 149, 905–915. doi: 10.1104/pp.108.130716

Singh, S., Brocker, C., Koppaka, V., Chen, Y., Jackson, B. C., Matsumoto, A., et al. (2013). Aldehyde dehydrogenases in cellular responses to oxidative/electrophilic stress. *Free Radic. Biol. Med.* 56, 89–101. doi: 10.1016/j.freeradbiomed.2012.11.010

Skibbe, D. S., Liu, F., Wen, T. J., Yandeau, M. D., Cui, X., Cao, J., et al. (2002). Characterization of the aldehyde dehydrogenase gene families of zea mays and arabidopsis. *Plant Mol. Biol.* 48, 751–764. doi: 10.1023/a:1014870429630

Székely, G., Abrahám, E., Cséplo, A., Rigó, G., Zsigmond, L., Csiszár, J., et al. (2008). Duplicated P5CS genes of arabidopsis play distinct roles in stress regulation and developmental control of proline biosynthesis. *Plant J.* 53, 11–28. doi: 10.1111/j.1365-313X.2007.03318.x

Tylichová, M., Kopecný, D., Moréra, S., Briozzo, P., Lenobel, R., Snégaroff, J., et al. (2010). Structural and functional characterization of plant aminoaldehyde dehydrogenase from pisum sativum with a broad specificity for natural and synthetic aminoaldehydes. *J. Mol. Biol.* 396, 870–882. doi: 10.1016/j.jmb.2009.12.015

Vasiliou, V., Pappa, A., and Petersen, D. R. (2000). Role of aldehyde dehydrogenases in endogenous and xenobiotic metabolism. *Chem. Biol. Interact.* 129, 1–19. doi: 10.1016/s0009-2797(00)00211-8

Wang, W., Jiang, W., Liu, J., Li, Y., Gai, J., and Li, Y. (2017). Genome-wide characterization of the aldehyde dehydrogenase gene superfamily in soybean and its potential role in drought stress response. *BMC Genomics* 18, 518. doi: 10.1186/s12864-017-3908-y

Wei, M. Y., Liu, J. Y., Li, H., Hu, W. J., Shen, Z. J., Qiao, F., et al. (2021). Proteomic analysis reveals the protective role of exogenous hydrogen sulfide against salt stress in rice seedlings. *Nitric. Oxide* 111-112, 14–30. doi: 10.1016/j.niox.2021.04.002

Wen, Y., Wang, X., Xiao, S., and Wang, Y. (2012). Ectopic expression of VpALDH2B4, a novel aldehyde dehydrogenase gene from Chinese wild grapevine (Vitis pseudoreticulata), enhances resistance to mildew pathogens and salt stress in arabidopsis. *Planta* 236, 525–539. doi: 10.1007/s00425-012-1624-z

Yang, J., An, D., and Zhang, P. (2011). Expression profiling of cassava storage roots reveals an active process of glycolysis/gluconeogenesis. *J. Integr. Plant Biol.* 53, 193–211. doi: 10.1111/j.1744-7909.2010.01018.x

Yang, Q., Han, X. M., Gu, J. K., Liu, Y. J., Yang, M. J., and Zeng. Functional, Q. Y. (2019). And structural profiles of GST gene family from three populus species reveal the sequence-function decoupling of orthologous genes. *New Phytol.* 221, 1060–1073. doi: 10.1111/nph.15430

Yang, H., Zhang, D., Li, H., Dong, L., and Lan, H. (2015). Ectopic overexpression of the aldehyde dehydrogenase ALDH21 from syntrichia caninervis in tobacco confers salt and drought stress tolerance. *Plant Physiol. Biochem.* 95, 83–91. doi: 10.1016/j.plaphy.2015.07.001

Yoshiba, Y., Kiyosue, T., Katagiri, T., Ueda, H., Mizoguchi, T., Yamaguchi-Shinozaki, K., et al. (1995). Correlation between the induction of a gene for delta 1-pyrroline-5-carboxylate synthetase and the accumulation of proline in arabidopsis thaliana under osmotic stress. *Plant J.* 7, 751–760. doi: 10.1046/j.1365-313x.1995.07050751.x

Yoshiba, Y., Kiyosue, T., Nakashima, K., Yamaguchi-Shinozaki, K., and Shinozaki, K. (1997). Regulation of levels of proline as an osmolyte in plants under water stress. *Plant Cell Physiol.* 38, 1095–1102. doi: 10.1093/oxfordjournals.pcp.a029093

Yoshiba, Y., Nanjo, T., Miura, S., Yamaguchi-Shinozaki, K., and Shinozaki, K. (1999). Stress-responsive and developmental regulation of Delta(1)-pyrroline-5-carboxylate synthetase 1 (P5CS1) gene expression in arabidopsis thaliana. *Biochem. Biophys. Res. Commun.* 261, 766–772. doi: 10.1006/bbrc.1999.1112

Yoshida, A., Rzhetsky, A., Hsu, L. C., and Chang, C. (1998). Human aldehyde dehydrogenase gene family. *Eur. J. Biochem.* 251, 549–557. doi: 10.1046/j.1432-1327.1998.2510549.x

Zhang, Y., Jiang, D., Yang, C., Deng, S., Lv, X., Chen, R., et al. (2021). The oxidative stress caused by atrazine in root exudation of pennisetum americanum (L.) k. schum. *Ecotoxicol Environ. Saf.* 211, 111943. doi: 10.1016/j.ecoenv.2021.111943

Zhang, Y., Mao, L., Wang, H., Brocker, C., Yin, X., Vasiliou, V., et al. (2012). Genome-wide identification and analysis of grape aldehyde dehydrogenase (ALDH) gene superfamily. *PloS One* 7, e32153. doi: 10.1371/journal.pone.0032153

Zhang, Q., Zhang, W. J., Yin, Z. G., Li, W. J., Zhao, H. H., Zhang, S., et al. (2020). Genome- and transcriptome-wide identification of C3Hs in common bean (Phaseolus vulgaris l.) and structural and expression-based analyses of their functions during the sprout stage under salt-stress conditions. *Front. Genet.* 11. doi: 10.3389/fgene.2020.564607

Zhou, S., Chen, X., Zhang, X., and Li, Y. (2008). Improved salt tolerance in tobacco plants by co-transformation of a betaine synthesis gene BADH and a vacuolar Na+/H+ antiporter gene SeNHX1. *Biotechnol. Lett.* 30, 369–376. doi: 10.1007/s10529-007-9548-6

Zhou, M. L., Zhang, Q., Zhou, M., Qi, L. P., Yang, X. B., Zhang, K. X., et al. (2012). Aldehyde dehydrogenase protein superfamily in maize. *Funct. Integr. Genomics* 12, 683–691. doi: 10.1007/s10142-012-0290-3

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 07 March 2023
DOI 10.3389/fpls.2023.1014349

OPEN ACCESS

EDITED BY
Jianghua Chen,
Key Laboratory of Tropical Plant Resource and Sustainable Use, Chinese Academy of Sciences (CAS), China

REVIEWED BY
Xiaoxu Li,
Chinese Academy of Agricultural Sciences (CAAS), China
Dayong Zhang,
Nanjing Agricultural University, China

*CORRESPONDENCE
Qiuying Zhang
zhangqiuying@iga.ac.cn

†These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION
This article was submitted to Functional and Applied Plant Genomics, a section of the journal Frontiers in Plant Science

RECEIVED 08 August 2022
ACCEPTED 20 February 2023
PUBLISHED 07 March 2023

CITATION
Chen H, Liu C, Li Y, Wang X, Pan X, Wang F and Zhang Q (2023) Developmental dynamic transcriptome and systematic analysis reveal the major genes underlying isoflavone accumulation in soybean. *Front. Plant Sci.* 14:1014349. doi: 10.3389/fpls.2023.1014349

Developmental dynamic transcriptome and systematic analysis reveal the major genes underlying isoflavone accumulation in soybean

Heng Chen[1,2,3†], Changkai Liu[1,2†], Yansheng Li[1,2], Xue Wang[1,2], Xiangwen Pan[1,2], Feifei Wang[1,2] and Qiuying Zhang[1,2*]

[1]Key Laboratory of Soybean Molecular Design and Breeding, Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Harbin, China, [2]Innovation Academy for Seed Design, Chinese Academy of Sciences, Harbin, China, [3]College of Advanced Agricultural Sciences, University of Chinese Academy of Sciences, Beijing, China

Introduction: Soy isoflavone, a class of polyphenolic compounds exclusively occurred in legumes, is an important bioactive compound for both plants and human beings. The outline of isoflavones biosynthesis pathway has been drawn up basically in the previous research. However, research on the subject has been mostly restricted to investigate the static regulation of isoflavone content in soybean, rather than characterize its dynamic variation and modulation network in developing seeds.

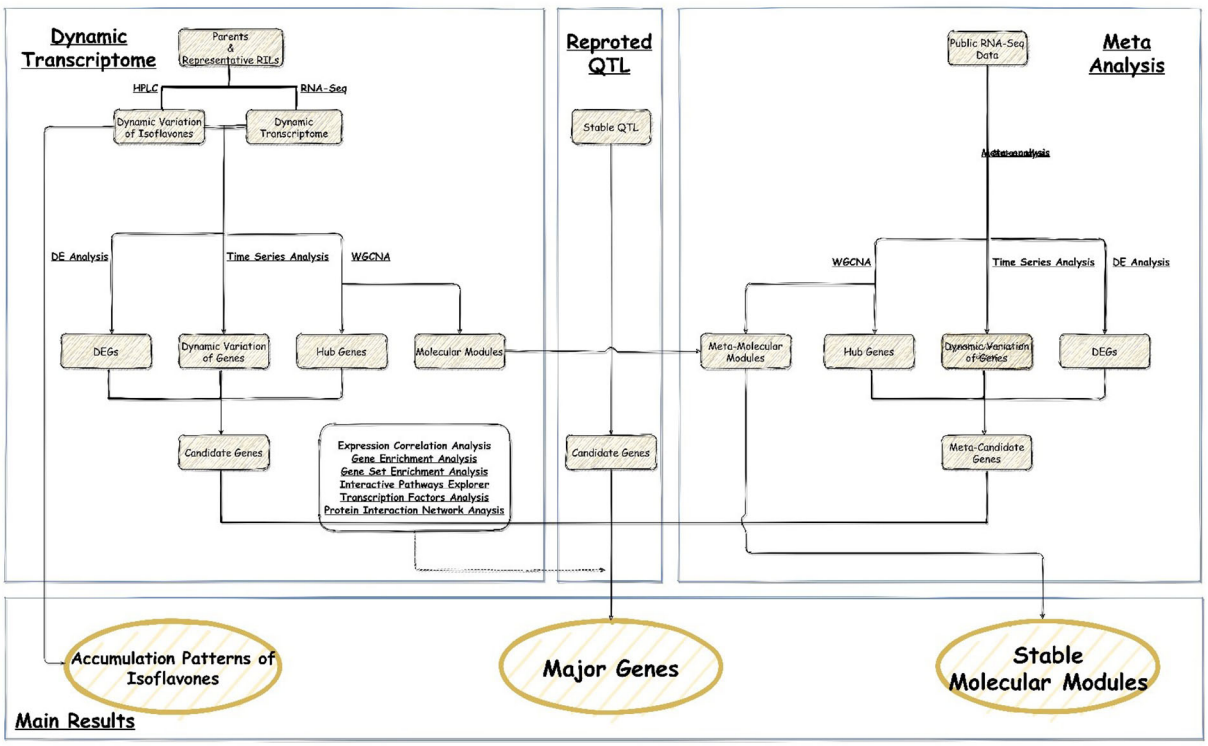
Methods: In this study, by using eight recombinant inbred lines (RIL), the contents of six isoflavone components in the different stages of developing soybean seeds were determined to characterize the dynamic variation of isoflavones, and the isoflavones accumulation pattern at physiological level was investigated. Meanwhile, we integrated and analyzed the whole genome expression profile of four lines and 42 meta-transcriptome data, based on the multiple algorithms.

Results: This study: 1) obtained 4 molecular modules strongly correlated with isoflavone accumulation; 2) identified 28 novel major genes that could affect the accumulation of isoflavones in developing seeds free from the limitation of environments; 3) discussed the dynamic molecular patterns regulating isoflavones accumulation in developing seed; 4) expanded the isoflavone biosynthesis pathway.

Discussion: The results not only promote the understandings on the biosynthesis and regulation of isoflavones at physiological and molecular level, but also facilitate to breed elite soybean cultivars with high isoflavone contents.

KEYWORDS

soy isoflavone, dynamic regulation networks, dynamic accumulation patterns, transcriptome, meta-analysis



GRAPHICAL ABSTRACT

1 Introduction

Soy isoflavone, a sub–group of flavonoids derived from phenylalanine, is a class of polyphenolic compounds exclusively occurred in legumes. This secondary metabolite is usually referred as phytoestrogen, since it is structurally and functionally similar to estrogen. Currently, isoflavones have been widely used in health product and bio–pharmaceutical, for it has the functions of preventing breast cancer, relieving Alzheimer disease, resisting menopause syndrome, et cetera (Dixon, 2004). Isoflavones also play a vital role in the system of plant defense responses. It could be further synthesized into major phytoalexins, and therefore exhibiting significant effect in plant insect–resistance, disease–resistance, and various other biotic stresses (Graham et al., 2007).

Because of the irreplaceable role of isoflavone in human health and plant growth, many related studies have been carried out, including its structure, ingredient, and genetic background. Up to now, a total of 12 kind of isoflavones are found, which are generally divided into 4 main categories based on their structure. They are: 1) aglycones (including daidzein, glycitein and genistein); 2) glycosides (including daidzin, glycitin and genistin); 3) acetyl–glycosides (including acetyl-daidzin, acetyl-glycitin and acetyl-genistin); 4) malonyl–glycosides (including malonyl-daidzin, malonyl-glycitin and malonyl-genistin) (Dixon, 2004). The glucosides are the hinges of isoflavone biosynthesis pathway *in vivo*, for they are the products of aglycones and the precursors of other two categories (Yu et al., 2003). Since the aglycones are the mainly active form in soybean seed while the glycosides are the primary storage form in soybean seed, the aglycones and glycosides are the most important concerned categories among these 12 isoflavones (Chen et al., 2021).

Isoflavone content is a typical quantitative trait, which is regulated by both genetic and environmental factors. The approaches of quantitative genetics have been the major methods for investigating the genetic background of isoflavone, such as QTL mapping. Based on the high-density genetic map, a total of 63 stable QTL (containing 52 meta-QTL and 11 novel QTL) were mapped by our proceeding study (Chen et al., 2021). However, although these QTL have been narrowed down to a short interval length, major genes influencing the accumulation of isoflavones are required for the quantitative trait genes (QTG), rather than QTL, regulate the phenotype *in vivo*.

Transcriptome analysis is a good way to obtain QTG from QTL, as well as an efficient and useful approach in exploring the major genes of complex quantitative trait (Stark et al., 2019). For instance, *CHS7* and *cytochrome P450* have been identified separately by Dhaubhadel et al. (2007) and (Guttikonda et al., 2010), which influenced the accumulation of isoflavone. Generally, candidate genes could be obtained from RNA–seq data by differential gene expression analysis (DE analysis). By comparing the transcriptional changes, the differential expression genes (DEG) among several significantly different samples could be obtained. However, the internal relationships between candidate genes could not be clarified in this

way. The weighted gene co–expression network analysis (WGCNA) to make up the inadequacy of DE analysis, for it could assist to describe the internal relationships among candidate genes by analyzing the degree of association between genes, finding molecular modules (clusters) and mapping hub genes (Zhao et al., 2010). The time series analysis is also applied to analyze RNA–seq data, which generally require a time series expression profile and could give the temporal expression variation of several important clusters or genes.

The outline of isoflavones biosynthesis pathway has been basically revealed for decades. However, the accumulation pattern in different developing stages of isoflavones at both molecular and physiological level is still unclear. Actually, there is a significant difference between the variation trend of different components, and a quite close internal relationship among different components of isoflavone (D'Agostina et al., 2008). The main factor influencing isoflavones variation might not be same in different developing stages. Consequently, clarifying the accumulation patterns of isoflavone components at physiological and molecular level is quite essential for comprehensively and deeply understanding the biosynthesis and regulation mechanisms of isoflavone in soybean.

The aims of current study were: 1) to investigate the accumulation patterns of isoflavones at physiological and molecular level; 2) to map the major genes and QTG associated with isoflavones accumulation; 3) to explore the dynamic regulation network regulating isoflavones accumulation in developing seed. The results could facilitate elucidating the molecular mechanism of isoflavones biosynthesis and regulation, as well as breeding elite soybean cultivars with high isoflavone content.

2 Materials and methods

2.1 Materials and field experiment

2.1.1 Materials for characterizing the phenotype dynamic variation

To construct the materials for characterizing the dynamic variation of six isoflavone components, the parental lines and six representative lines of a RIL population (Figure 1) constructed by our research group were adopted (Chen et al., 2021). Briefly, by using the single–seed descendent method, a F7 population named as ZH RIL with 162 lines was obtained based on a pair of parents, 'Zhongdou 27' (ZD27) and 'Hefeng 25' (HF25).

The two cultivars exhibit significant difference at phenotype level: 'Zhongdou 27' containing high isoflavone contents, a late–maturing cultivar, mainly planted in Huang–Huai–Hai region of China (Northern China); while 'Hefeng 25' containing low isoflavone contents, a higher–yielding cultivar, is mainly planted in Northeast China. Meanwhile, the six selected lines are also significantly different from each other at phenotypic level, as shown by the results of isoflavone determination (Figure 1). All of these materials were planted in the Acheng district, Harbin, Heilongjiang province, China (45.6°N, 126.6°E) in 2021. The main materials information could be found in Figure 1 and Table 1.

For characterizing the complete process of isoflavone accumulation during seed development, we selected 10 representative periods (S1-S10) in this part (Figure 1 and Table 1). Samples at S1 period were collected at R5 of soybean developing stage, while samples at S10 period were at full maturity stage. Each interval period was 7 days.

2.1.2 Materials for RNA-seq and q-RTPCR

To construct the materials for RNA–seq and qRT–PCR analysis, the parents (ZD27 and HF25) and two representative lines (ZH01 and ZH02) of the RIL population were adopted.

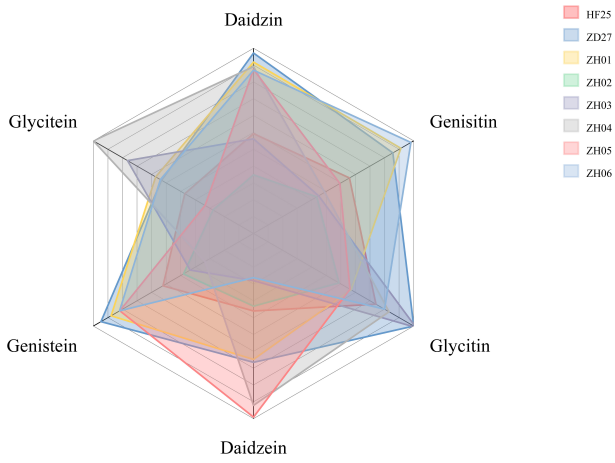


FIGURE 1
The isoflavone content of samples in this study (The distance between central point and vertex of each color block represents the content of isoflavone component in a sample).

TABLE 1 Basic information of filed experiments.

Experiments	Samples	Periods	Field experiment locations
Dynamic variation of isoflavones	ZD27 HF25 ZH01 ZH02 ZH03 ZH04 ZH05 ZH06	S1-S10	Acheng district, Harbin
RNA-seq	ZD27 HF25 ZH01 ZH02	S2-S8	Acheng district, Harbin
qRT-PCR	ZD27 HF25 ZH01 ZH02	S2-S8	Acheng district, Harbin

They were planted in the Acheng district, Harbin, Heilongjiang province, China. According to the time series, 7 groups (S2-S8) of soybean seeds were used for RNA-seq (Figure 1 and Table 1).

2.1.3 Field experiment and sample collection

These materials were sown in rows with 3 m long, 0.65 m wide and a distance with 0.08 m between the individual plants. Field management followed normal soybean production practices. The main information of the field experiment design could be found in Figure 1 and Table 1.

Soybean seeds for RNA-seq were collected into pre–cooling and RNase–free cryogenic vials (Biosharp, Anhui, China), and then quickly frozen and stored at -80 °C condition. Soybean seeds for isoflavone determination were treated by the following steps: 1) green removed at 105 °C for 30 mins; 2) dried at 45 °C for 48 hours; 3) stored at -20 °C condition.

2.2 Phenotype identification

To determine the contents of six isoflavones in soybean seed, the high–performance liquid chromatography (HPLC) method was utilized, which is one of the most effective and common approach for isolating and determining complex metabolites. The detailed operations and conditions of HPLC assay were described in our previous paper (Chen et al., 2021).

2.3 RNA-seq and quality control

To obtain the whole genome expression profile of the developing seeds, we performed the following steps in turn: total RNA isolation, RNA sequencing and *de novo* transcriptome assembly. Several kits, programs and instruments were also utilized, whose detailed operations could be acquired on the tutorials or instructions of manufacturer.

For the isolation of total RNA, the Ultrapure RNA Kit (ComWin Biotech, Beijing, China) was used. The extracting solution of total RNA was stored at -80 °C condition. RNA quality was checked and evaluated by the Agilent 2100 Nano Bioanalyzer (Agilent Technologies, California, USA) and NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Massachusetts, USA). The total RNA extraction was permitted to construct cDNA library, if and only if it could meet the criteria as follow: 1) RNA content ≥ 1.0 μg and RNA concentration ≥ 35 ng/μl; 2) OD260/OD230 ≥ 1.0 and OD260/280 ≥ 1.8; 3) RNA Integrity Number (RIN) ≥ 7.

For RNA sequencing, the Illumina Truseq™ RNA sample prep Kit (NEB, Massachusetts, USA) and Illumina NovaSeq 6000 platform (Illumina, California, USA) was used to perform RNA paired–end (PE) sequencing. Then, the BCL convert (Illumina, California, USA) was used to transform the sequence results to raw reads. The Phred Score (QPhred) and base distribution were selected to evaluate the quality of RNA sequencing. The formula of Phred Score is:

$$QPhred = -10 \quad \log_{10} e$$

where the e represents the error rate of base sequencing.

For *de novo* transcriptome assembly, StringTie2 (Kovaka et al., 2019) was utilized. The gffcompare, a subassembly of StringTie2, was also used to discover new transcripts, genes and/or exons based on the reference genome Williams 82. a4. v1 (Schmutz et al., 2010) from Phytozome v13 (Goodstein et al., 2011).

2.4 Transcriptome analysis

To further analyze the whole genome expression profile of developing seeds, multiple algorithms were applied to characterize the clusters and genes associated with isoflavone accumulation. Several sorts of software were involved, whose specific operation could be found in their tutorials or references.

2.4.1 Whole genome expression quantification and analysis

For quantifying the whole genome expression, the program Salmon (Patro et al., 2017) was used. The transcripts per million reads (TPM) was selected as standards to evaluate the expression level, whose formula is as follows:

$$TPM = \frac{R \times 10^6}{(\frac{R_1}{l_1} + \frac{R_2}{l_2} + \cdots + \frac{R_n}{l_n} \times l)}$$

where the R and l represents the read counts and length of the analyzed gene, the R_1 to R_n represents the read counts of the *1st* gene to *nth* gene, the l_1 to l_n represents the gene length of *1st* gene to *nth* gene.

For analyzing the variation of whole genome expression between different biological repetitions of a single sample, we performed the correlation analysis with Pearson correlation coefficient. The result provides a reference for DE analysis as well. Additionally, for characterizing and investigating the variation of whole genome expression among different samples, three algorithms were utilized, including Venn analysis and correlation analysis (with Pearson correlation coefficient).

2.4.2 Selected algorithms

For detecting the DEGs among different samples, DE analysis was performed with the raw read counts *via* the R package DEseq2 (Love et al., 2014). The parameters set for this algorithm were *FDR q – val< 0.05* and $|\log_2 FC| \geq 1$ (fold change). Moreover, for obtaining the gene clusters and hub genes correlated with isoflavones accumulation, the R package WGCNA (Zhang and Horvath, 2005) was utilized.

For analyzing the dynamic variation of transcript profile, two programs, the maSigPro (Nueda et al., 2014) and Short Time–series Expression Miner (STEM) (Ernst and Bar-Joseph, 2006), were respectively applied to the temporal expression differences and temporal expression trends during seeds development. In this process, we selected 6 reported genes as standard to screen major

genes of isoflavone accumulation. Their expression variation in developing seeds was characterized based on RNA-seq data of this study (Figure S1).

For further analysis and selection of the candidate genes based on their properties and functions, several other databases and related procedures were also applied, including Plant Transcription Factor Database V5.0 (PlantTFDB) (Tian et al., 2019) and STRING V11 (Szklarczyk et al., 2018).

For annotating and analyzing the functions of candidate genes, the gene functional annotation and enrichment analysis were used in the study. In this process, the R package ClusterProfiler (Yu et al., 2012) was used in KEEG and GO enrichment analysis, which was based on the Fisher's precision probability test with $P - adjust > 0.05$.

2.5 Meta-analysis

To collect and integrate the RNA-seq data from the previous studies, the meta–analysis (Borenstein et al., 2009) was used. The basic information of meta-data collected in this study were: cultivars' name, period of seed development, experiment time, experimental conditions, sequencing techniques, sequencing platform, library layout, treatment and source (accession number and reference).

Approximately 700 Gbp meta-data were collected from the National Center of Biotechnology Information (NCBI) and National Genomics Data Center (NGDC), including whole genome expression profiles of 80 developing soybean seeds. The transcriptome data was divided into 4 groups, whose main information was as shown in Table 2 and Table S1. To be specific, the materials of group 1 were formed by 'Suinong 14' (SN14) and its four (chromosome segment substitution lines) CSSLs: SN14a, SN14b, SN14c and SN14d. Each germplasm contained samples with 3 stages: S1 (early maturation), S2 (middle maturation) and S3 (dry seed). The samples of group 2 consisted of 'Williams 82' (Wm82). Each cultivar in group 2 contained samples with 4 stages: S1 (early maturation), S2 (middle maturation), S3 (late maturation) and S4 (dry seed). The samples of group 3 included Dongnong 47 (DN47) and its near–isogenic line (DN47n), and each cultivar comprised seeds with 5 stages (days after flowering, DAF): S1 (18 DAF), S2 (25 DAF), S3 (35 DAF), S4 (50 DAF) and S5 (55 DAF). The group 4 was formed by 2 RIL derived from V99–5089 and CX–1834, VC–M and VC–m, and these materials involved 5 periods: S1 (2 – 4 mm seed), S2 (4 – 6 mm seed), S3 (6 – 8 mm seed), S4 (8 – 10 mm seed) and S5 (10 – 12 mm seed).

The collected raw data were treated by the following steps: 1) quantifying the data by using Salmon software and TPM algorithm; 2) performing multiple algorithms in transcriptome analysis, including DE analysis, WGCNA and time series analysis; 3) performing gene functional annotation and enrichment analysis. For simplifying the analysis process in this part, the following conditions and requirements were set: 1) the Williams 82.a4.v1 was selected as the reference genome; 2) the expression abundance of *IFS1* (*Glyma.07G202300*) and *CHS7* (*Glyma.01G228700*) were used to define the strength of isoflavone accumulation; 3) the top 5000 genes with maximum mean absolute deviation (MAD) of each meta-data group were used to perform WGCNA; 4) the gene cluster generated by WGCNA were selected with the combination with metabolic pathway analysis by iPath; 5) the gene clusters number of time series analysis was set as nine.

2.6 Linking candidate genes with isoflavone-related QTL

To further screen the candidate genes, the association analysis between the candidate genes and QTL was performed. Based on the 63 stable QTL (52 meta-QTL and 11 novel QTL) mapped in our previous study (Table S2) (Chen et al., 2021), the genes acquired from the transcriptome analysis and meta-analysis were selected through the co-localization with these QTL to obtain QTG underlying the accumulation of isoflavone. The results could further understand the relationship between the candidate genes and isoflavone components.

2.7 qRT-PCR

To validate the reliability of RNA–seq and select the major genes obtained from transcriptome analysis and meta–analysis, the

TABLE 2 The basic information of Meta–data collected in this study.

Group	Germplasms	Number of Sample	Meta–data Size	Platform	Accession and Reference
G1	SN14 SN14a SN14b SN14c SN14d	45	355 Gbp	Illumina HiSeq 4000	PRJCA000523 (Qi et al., 2018)
G2	Wm82	12	9 Gbp	Illumina HiSeq 2000	PRJNA388955 (Pelletier et al., 2017)
G3	DN47 DN47n	20	211 Gbp	Illumina Hiseq 2000	PRJNA315512 (Song et al., 2016)
G4	VC–M VC–m	30	118 Gbp	Illumina Hiseq 2000	PRJNA304631 (Redekar et al., 2015)

qRT–PCR analysis was conducted based on a LightCycler 480II platform (Roche, Basel, Switzerland) with MagicSYBR Mixture (ComWin Biotech, Beijing, China). The total RNA was used to construct cDNA library immediately by HiFiScript cDNA Synthesis Kit (ComWin Biotech, Beijing, China)., and then stored at -20 °C condition. The primers were listed in Table S3, and the PCR amplification conditions were set according to Table S4. The internal control was *GmUKN1* (*Glyma.12G020500*) (Hu et al., 2009). Three independent experiments were executed for each sample. The relative abundance of gene expression was estimated *via* the underlying comparative threshold method ($2^{-\Delta\Delta CT}$) (Pfaffl, 2001).

3 Results

3.1 Dynamic variation of isoflavone components

The accumulation patterns and overall trends of each isoflavone contents in developing seeds were summarized in Figure 2.

The accumulation pattern of daidzin is "climbing mode": rising gently and then reaching equilibrium. Specifically, the accumulation of daidzin started during S1-S2, increased rapidly at S4, and reached equilibrium at S7 or S8.

The accumulation pattern of daidzein is "two peaks mode", which could be divided into 3 phases: 1) rising and then decreasing (S1-S5); 2) rising and decreasing again (S5-S8); 3) reaching equilibrium (S9-S10). The first peak value was occurred at S2 or S3, while the second peak value was usually occurred at S7 or S8.

The accumulation pattern of glycitin is "single peak and climbing mode", which is similar to pattern of daidzin accumulation but have a peak at S5. The glycitin contents reached its equilibrium at S6 or S7 (except ZH02 which reached its equilibrium at S3), which was notably earlier than other isoflavone components.

The accumulation pattern of glycitein is "single peak mode": 1) increased during S1-S4; 2) decreased slowly during S4-S7; 3) kept constant till the end. Of course, there were some exceptions. The variation of glyceitin contents in ZD27 started at S2 and ended at S5, which exhibited a short and early variation process; while ZH04 had a long and late accumulation process which reached the peak at S7 and ended at S8.

The accumulation pattern of genistin is "two-part climbing mode". In the first part (S1-S4): the contents of genistin were unchanged during S1-S3, and then increased slowly during S3-S4. In the second part (S5-S10): genistin contents increased at a high rate during S5-S7, and stopped increasing and maintained constant during S7-S9, and declined lightly during S9-S10. Two significant characteristics were suggested: 1) the difference of genistin contents among different lines firstly appeared at S4; 2) the genistin contents of matured seed were depended on the genistin biosynthesis rate during S5-S7.

The accumulation pattern of genistein is "bipolar peak mode", which exhibits a special variation model compared with

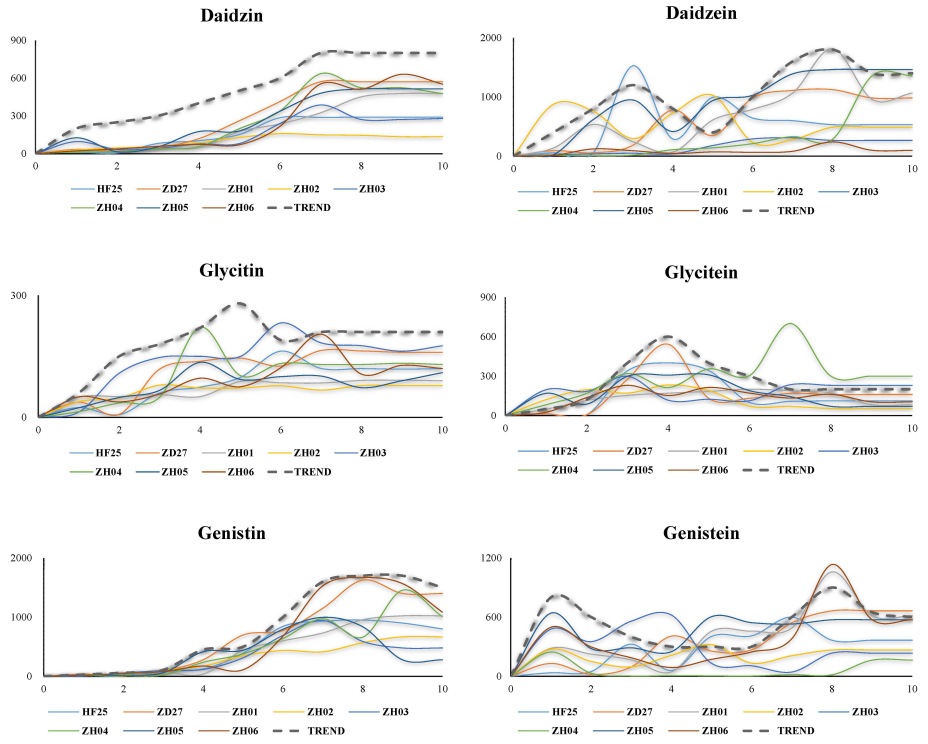


FIGURE 2
Dynamic variation of six isoflavone compounds in developing soybean seed (Each broken line represents the variation trend of isoflavone content, while each full line represents the real situation of the change of isoflavone content).

other isoflavones. The pattern could be separated to two parts: 1) S1-S5 (the first peak appeared at S1 or S2, and then decreased till S5); 2) S5-S10 (the genistein restarted to accumulate at S5, and reached second peak at S8, and then decreased lightly and kept constant).

3.2 Transcriptome analysis

3.2.1 RNA–seq evaluation

Based on the Illumina NovaSeq 6000 platform, this study obtained 444.5G clean reads (with a Q30 ratio of more than 90%) from the 56 soybean germplasms (Table S5). The results of quality control suggested a high quality of sequencing data (Table S5 and Figure S2A), and the comparative analysis of data indicated that the samples of RNA-seq were highly pure and the reads were appropriate to gene assembly (Figure S2B). Evaluating the quality of sequencing data indicated that: 1) the gene annotation is quite complete (Figure S2C); 2) the coverage of sequencing was great and homogeneous (Figure S2D). Therefore, the high-quality RNA–seq data were obtained.

By transcript assembly and gene functional annotation, the current research obtained: 1) 118347 transcripts and 50556 genes (Table S6). The expression amount of these genes and transcripts were quantified by Salmon (Figure S2E) (Patro et al., 2017). This indicates that the samples and their transcriptome data could characterize the development of soybean seeds in whole. Several specifically expressed and co-expressed genes were acquired, by the Venn analysis among the different stages of the same germplasm and among the different germplasms in the common stage (Figure 3). Moreover, the expression amounts of genes exhibited consistency with qRT-PCR results of selected genes in this study, indicating a strong reliability of RNA-seq results (Figure S3). All in all, the transcript profile of this research is strongly reliable, repeatable and representative, as well as suitable for further analysis.

3.2.2 DE analysis

From the perspective of germplasm and period, the RNA-seq data of 56 soybean samples were divided into 11 groups (4 germplasm groups and 7 period groups) for DE analysis (Figure S4). Among them, 11 times of DE analysis were carried out for each germplasm group, and 6 times of DE analysis were carried out for each period group.

In ZH01 group, 36571 DEGs were detected, including 17228 up-regulated genes and 19343 down-regulated genes. In ZH02 group, 76744 DEGs were generated, including 36051 up-regulated genes and 40693 down-regulated genes. In ZD27 group, 36773 DEGs were detected, including 16679 up-regulated genes and 20094 down-regulated genes. In HF25 group, approximately 41000 DEGs were obtained, including about 20000 up-regulated genes and 21000 down-regulated genes (Figure S4). Additionally, we also detected 4286 DEGs in S2 group, 3231 DEGs in S3 group, 5536 DEGs in S4 group, 11542 DEGs in S5 group, 36 DEGs in S6 group, 12338 DEGs in S7 group and 1999 DEGs in S8 group (Figure S4).

3.2.3 DEGs related to the dynamic variation of isoflavone components

According to the variation trend of daidzin content in the developing seeds, the research suggests that the related main genes should be significantly differential expression in S3-S4 (the beginning of daidzin accumulation) and S6-S7 (the end of daidzin accumulation) period, which may be slightly different in various lines. Thus, the DEGs of 8 comparison groups were selected to examine the genes modulating daidzin accumulation, as shown in Table S7. The results of gene functional annotation showed that 122 genes may be closely related to the accumulation of daidzin. Comparing the expression of these 122 genes in different samples at S4 (the daidzin content began to differ among each sample) and S7 (the daidzin is close to the end of accumulation in ZD27, and is still accumulating in ZH01, but stops accumulating in HF25 and ZH02). This indicates that the 38 genes of S4 stage should play an

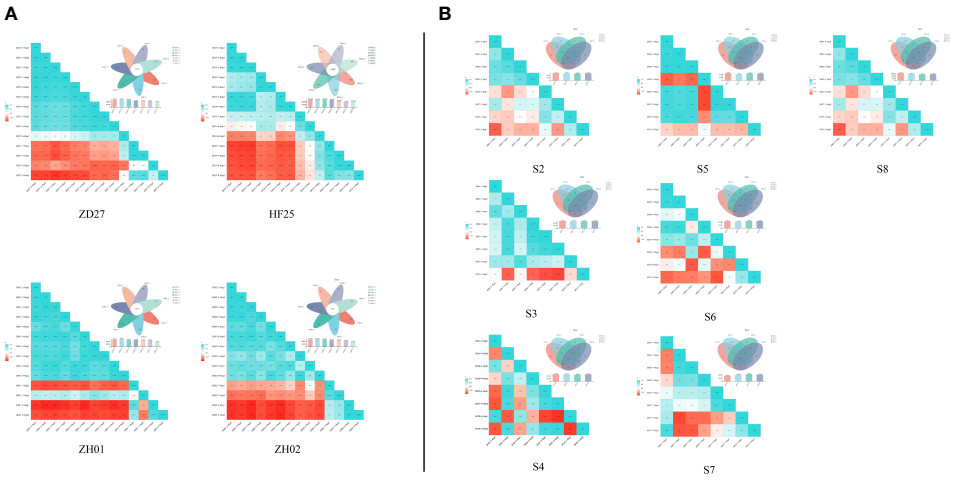


FIGURE 3
Venn analysis and correlation analysis among the RNA-seq of different samples **(A)** Venn analysis and correlation analysis of horizontal comparison; **(B)** Venn analysis and correlation analysis of lengthways comparison.

important role in the initial stage of daidzin accumulation, while the 61 genes of S7 stage should influence the terminal stage of daidzin accumulation (Table S7). A total of 23 genes functionally related to daidzin accumulation were significantly differentially expressed in both initial and terminal stage, indicating that these genes might be involved in the whole process of daidzin accumulation. Briefly, the above results suggest that: 1) 15 genes significantly affect the initiation of daidzin accumulation; 2) 38 genes are involved in the termination of daidzin accumulation; 3) 23 genes are involved in the whole process of daidzin accumulation.

Based on the variation trend of daidzein content, the current study suggests that the daidzein related major genes are significantly differentially expressed in S2-S3 (the rising section of the first peak of daidzein accumulation), S3-S4 (the falling section of the first peak of daidzin accumulation), S6-S7 (the rising section of the second peak of daidzein accumulation) and S7-S8 period (the end of daidzein accumulation). According to the above characteristics, the DEGs of 19 comparison groups were selected (Table S7). With the combination of gene functional annotation and gene enrichment analysis, a total of 198 DEGs were obtained. *Via* further comparing these genes with the DEGs obtained from lengthways comparison (S2, S4, S6 and S7 were selected), we found 132 candidate genes involved in the accumulation of daidzein (Table S7), including: 1) 38 genes significantly affecting the initial accumulation of daidzein at the first stage; 2) 59 genes involved in the termination of daidzein accumulation at first stage; 3) 20 genes involved in the whole process of daidzein accumulation at first stage; 4) 48 genes significantly affecting the initiation of daidzein accumulation at the second stage; 5) 79 genes involved in the termination of daidzein accumulation at the second stage; 6) 24 genes involved in the whole process of daidzein accumulation at the second stage; 7) 10 genes involved in the whole process of daidzein accumulation.

Combining with the change trend of glycitin content and the results of DE analysis, the glycitin related genes were significantly differentially expressed in S2-S3 (rising stage of glycitin accumulation) and S5-S6 (falling stage of glycitin accumulation), with slight difference in various soybean germplasms. From the above perspective, DEGs of 6 comparison groups were used for researching the major genes of glycitin accumulation (Table S7). Screening with gene functional annotation, 155 genes were obtained. The differential expression of these genes in different materials at the same period (S2, S4 and S5 were selected) indicates that 119 genes are involved in the accumulation of glycin during the development of soybean seed (Table S7), including: 1) 36 genes involved in the beginning of the increase stage of glycin content; 2) 44 genes involved in the termination of increase stage and the beginning of decrease stage; 3) 106 genes involved in the termination of the decline stage; 4) 17 genes involved in the whole increasing process; 5) 36 genes involved in the whole decline process; 6) 16 genes involved in the whole process of glycitin accumulation.

Due to the dynamic variation of glycitein content is "single peak mode", we argue that the major genes of glycitein accumulation should be signally varied in S2-S4 (rising stage of glycitein accumulation) and S4-S7 (falling stage of glycitein accumulation). Consequently, DEGs from 9 lengthways comparison groups were utilized to screen candidate genes (Table S7). The results manifest that 156 genes are closely associated with the accumulation of glycitein. *Via* further comparing these genes with the DEGs obtained from lengthways comparison (S2, S4 and S5 were selected), 138 genes were found (Table S7), containing: 1) 34 genes participated in the beginning of increase stage of glycitein content; 2) 46 genes participated in the termination of increase stage and the beginning of decrease stage; 3) 130 genes participated in the termination of decline stage; 4) 19 genes participated in the whole increase process; 5) 43 genes participated in the whole decline process; 6) 18 genes participated in the whole process of glycitein accumulation.

For the variation trend of genistin content is "two-part climbing mode", the expression of genistin related major genes should be markedly different in S3-S4 (the first stage of increase of genistin content), S5-S6 (the second stage of the increase) and S6-S7 (the cessation of genistin accumulation). Based on this view, DEGs of 11 lengthways comparison groups were chosen for the investigation of candidate genes (Table S7). With gene functional annotation, 187 genes of genistin were attained. By comparing with the DEGs obtained from lengthways comparison (S3, S4, S5 and S7 were selected), 170 genes involved in the accumulation of genistin during the development of soybean were attained (Table S7). These candidate genes are consisted of: 1) 47 genes associated with the beginning of the first stage of genistin content increase; 2) 56 genes associated with the termination of the first stage; 3) 24 genes associated with the whole process of first stage; 4) 56 genes associated with the beginning of the second stage; 5) 90 genes associated with the termination of the second stage; 6) 63 genes associated with the whole process of the second stage; 7) 15 genes involved in the whole process of genistin accumulation.

Since the accumulation pattern of genistein content is "bipolar peak mode", the differential expression of major genes related to genistein accumulation was found in S0-S1 (the rising part of the first peak of genistein content), S1-S2 (the falling part of the first peak), S5-S6 (the rising part of the second peak) and S8-S9 (the falling part of the second peak). The DEGs of 22 lengthways comparison groups were selected to investigate the genes modulating genistein accumulation (Table S7). A total of 212 genes were found. Further comparing with the DEGs of 4 lengthways comparison groups (S2, S4, S6 and S8), 83 candidate genes were selected (Table S7), containing: 1) 38 genes significantly affecting the beginning of the first stage of genistein accumulation; 2) 52 genes affecting the termination of the first stage; 3) 19 genes affecting the whole process of the first stage; 4) 45 genes affecting the initiation of the second stage; 5) 43 genes affecting the termination of the second stage; 6) 10 genes affecting the whole process of second stage; 7) only the gene, *Glyma.13G206700* (encoding the glucose-6-phosphate transporter), that affects the whole process of genistein accumulation.

Integrating all candidate genes, we found 60 genes that affect the accumulation of both daidzin and daidzein, 80 genes that affect the accumulation of both glycitin and glycitein, and 64 genes that affect the accumulation of genistein and genistein. Additionally, the results also showed that 25 genes regulated the accumulation of

all components of glucoside isoflavones, while 39 genes modulated the accumulation of all components of aglycone isoflavones.

To sum up, 250 DEGs related to isoflavone accumulation were obtained in this study, including 76 genes influencing the accumulation of daidzin, 132 genes influencing the accumulation of daidzein, 119 genes influencing the accumulation of glycitin, 138 genes influencing the accumulation of glycitin, 170 genes influencing the accumulation of genistein, and 83 genes influencing the accumulation of genistein (Table S7).

3.2.4 WGCNA

Based on the phenotypic and transcriptome data, the WGCNA was performed to analyze the internal relationship of isoflavone related genes. This study constructed 11 molecular modules (Figures S5A, S5B). Combining with the results of correlation analysis (Figure S5C), module significance analysis (MS) (Figures S5D, S5E) and iPath (Figure S5F), three modules (yellow, purple and turquoise) were suggested to play important roles in the biosynthesis of isoflavones. Additionally, although the brown module was judged as insignificant in correlation analysis, we still proposed that the brown gene module also plays an important role in isoflavones accumulation. Because: 1) the brown module consists of lots of candidate genes with high gene significance (GS) value; 2) the MS value of brown module is high; 3) according to the cluster analysis and correlation analysis (Figure S5C), the brown module is strongly correlated with yellow and turquoise module. Consequently, the four gene modules (Table S8), yellow, purple, turquoise and brown, should have vital functions in the accumulation of isoflavones in developing soybean seeds.

According to the GS analysis, 434 and 477 candidate genes were highly correlated with the glucosides and aglycone isoflavones respectively (GS $\geq$ 0.5), involving yellow, brown and purple molecular modules. Some candidate genes have been identified to participate in the accumulation of isoflavones, such as *Glyma.01G228700* (encoding chalcone synthase), *Glyma.07G202300* (encoding 2-hydroxyisoflavone synthase) and *Glyma.13G173401* (encoding Cytochrome P450 monooxygenases) (Dhaubhadel et al., 2003; Dhaubhadel et al., 2007; Guttikonda et al., 2010), which indirectly supported that the candidate genes detected based on WGCNA algorithm were highly credible.

In brief, the 4 gene clusters, yellow, purple, turquoise and brown (Table S8), were identified to have a great impact on the accumulation of isoflavones in the developing seeds of soybean. A total of 828 candidate genes with high GS value were also obtained, including 434 and 477 candidate genes correlated with glucosides and aglycones respectively.

3.2.5 Time series analysis and comprehensive analysis

Thirty gene clusters were generated by time series analysis (Figure S6). The expression trend of genes of 24 clusters was close to that of genes promoting isoflavone accumulation, including cluster 1, cluster 2, cluster 3, cluster 6, cluster 9, cluster 10, cluster 11, cluster 12, cluster 13, cluster 14, cluster 15, cluster 16, cluster 17, cluster 18, cluster 19, cluster 20, cluster 21, cluster 22, cluster 23, cluster 25, cluster 26, cluster 27, cluster 29 and cluster 30. While the expression trend of genes of cluster 4, cluster 5, cluster 7, cluster 8, cluster 24 and cluster 28 was close to that of genes inhibiting isoflavone accumulation. Integrating the dynamic change of isoflavone content (Figure 2), the expression abundance variation of reported genes (Figure S1) and the results of gene function annotation, a total of 362 genes were given in this part, which may play an important role in the accumulation of isoflavones.

Finally, further Integrating with the results of DE analysis, WGCNA and time series analysis, 210 candidate genes were suggested to have a significant impact on the accumulation of isoflavones during soybean seed development.

3.3 Meta-analysis

3.3.1 G1

The materials of G1 were divided into 5 categories: SN14, SN14a, SN14b, SN14c and SN14d. To analyze the whole genome expression profile of these samples, DE analysis (Figure S7A), WGCNA (Figure S7B) and times series analysis (Figure S7C) were performed. Initially, a total of 452 DEGs, functionally associated with the accumulation of isoflavone and its components, was obtained. Additionally, 241 hub genes, involved 12 molecular modules, were generated. With the functional annotation, 3 gene modules and 23 hub genes were mapped in this study (Table S9). Meanwhile, according to the results of time series analysis, the expression tendency of seven gene clusters (cluster 1, cluster 3, cluster 4, cluster 5, cluster 6 and cluster 7) was in accordance to the expression patterns of genes which promoted the isoflavone accumulation. The expression tendency of cluster 2 and cluster 9 seemed like the expression patterns of genes which suppressed the isoflavone accumulation.

3.3.2 G2

The materials of G2 contained 4 developing stages of Wm82. To analyze the whole genome expression profile of these samples, DE analysis (Figure S8A), WGCNA (Figure S8B) and times series analysis (Figure S8C) were performed. Initially, a total of 362 DEGs, functionally associated with the accumulation of isoflavone and its components, were obtained. Additionally, 141 hub genes, involved 8 molecular modules, were generated. With the functional annotation, 3 gene modules and 16 hub genes were mapped in this study (Table S9). Meanwhile, according to the results of time series analysis, the expression tendency of four gene clusters (cluster 2, cluster 3, cluster 4 and cluster 7) was in accordance to the expression patterns of genes which promoted the isoflavone accumulation. The expression tendency of cluster 1, cluster 5, cluster 6, and cluster 9 seemed like the expression patterns of genes which suppressed the isoflavone accumulation.

3.3.3 G3

The materials of G3 were divided into DN14 and DN14n. To analyze the whole genome expression profile of these samples, DE analysis (Figure S9A), WGCNA (Figure S9B) and times series

analysis (Figure S9C) were performed. Initially, a total of 563 DEGs, functionally associated with the accumulation of isoflavone and its components, were obtained. Additionally, 161 hub genes, involved 7 molecular modules, were generated. With the functional annotation, 3 gene modules and 23 hub genes were mapped in this study (Table S9). Meanwhile, according to the results of time series analysis, the expression tendency of four gene clusters (cluster4, cluster 6, cluster 7 and cluster 8) was in accordance to the expression patterns of genes which promoted the isoflavone accumulation. The expression tendency of 4 clusters, including cluster 2, cluster 3, cluster 8 and cluster 9, seemed like the expression patterns of genes which suppressed the isoflavone accumulation.

3.3.4 G4

The materials of G4 were divided into VCM and VCm. To analyze the whole genome expression profile of these samples, DE analysis (Figure S10A), WGCNA (Figures S10B, C) and times series analysis (Figure S10D) were performed. Initially, a total of 642 DEGs, functionally associated with the accumulation of isoflavone and its components, were obtained. Additionally, the RNA-seq data of VCM and VCm were used to perform WGCNA respectively, for there were significant difference between these two germplasms. A total of 121 hub genes, involved 6 molecular modules, were generated *via* analyzing the RNA-seq data of VCM; while 141 hub genes and 7 molecular modules were obtained *via* performing WGCNA with transcriptome profile of VCm. With the functional annotation, 3 gene modules and 13 hub genes were mapped in this study (Table S9). Meanwhile, according to the results of time series analysis, the expression tendency of seven gene clusters (cluster 2, cluster 3, cluster 5, cluster 6, cluster 7, cluster 8 and cluster 9) was in accordance to the expression patterns of genes which promoted the isoflavone accumulation. The expression tendency of cluster 1 and cluster 4 seemed like the expression patterns of genes which suppressed the isoflavone accumulation.

Briefly, a total of 1436 DEGs, 54 hub genes and 9 molecular modules (Table S9) were attained by summarizing the results of meta-analysis. With further combination of gene functional annotation and gene enrichment analysis, 292 genes which might affect the accumulation of isoflavones were acquired.

3.4 Stable major genes and molecular modules of isoflavone accumulation

The stable major genes and molecular modules of isoflavone accumulation were suggested in this study *via* the following steps: 1) integrating the results of transcriptome analysis and meta-analysis; 2) linking candidate genes with isoflavone-related QTL.

3.4.1 Integrating the results of transcriptome analysis and meta-analysis

Among the 210 and 292 candidate genes obtained from the transcriptome analysis and meta-analysis respectively, 41 genes were overlapped. Therefore, these genes are stable major genes which could affect the accumulation of isoflavones free from the limitation of environments and germplasms. (Table S10 and Figure 4A).

Additionally, we obtained 4 gene modules and 9 meta-molecular modules *via* WGCNA with RNA-seq data and meta-data respectively. Comparing these modules, we found that each gene module could find at least 2 similar meta-molecular modules (some genes were overlapped). For instance, in the brown module: 1) 752 genes were similar to the blue module of G2 group meta-data; 2) 614 genes were similar to the blue module of G3 group meta-data; 3) 1195 genes were similar to the turquoise module of G4-VCM group meta-data; 4) 635 genes were similar to the blue module of G4-VCm group meta-data. This suggests that: 1) the internal effects and correlation of all molecular modules are rather

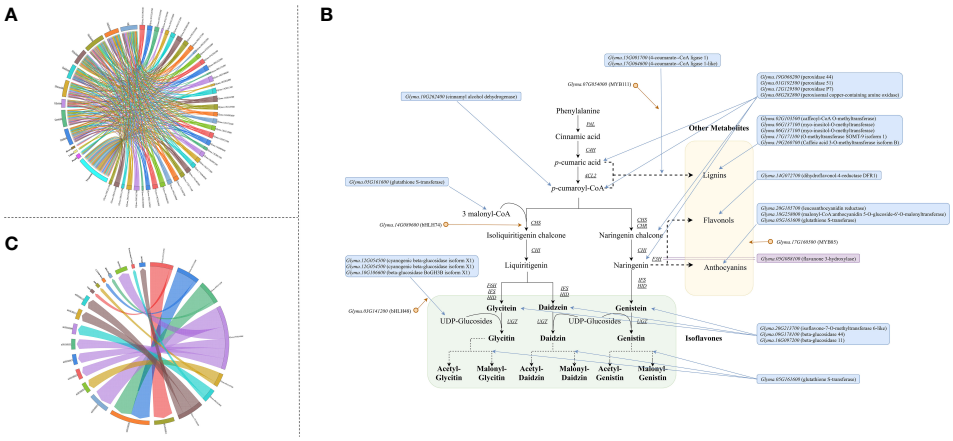


FIGURE 4
Major genes underlying isoflavone accumulation free from the limitation of environments **(A)** Major genes underlying isoflavone accumulation and their effects (The underlined words are isoflavone components, and the italicized words are QTL name, and the bold words are gene module generated by WCGNA); **(B)** QTG underlying isoflavone accumulation and corresponded stable QTL (The underlined words are isoflavone components, and the italicized words are gene name, and the bold words are gene annotation description); **(C)** Expansion of isoflavone biosynthesis pathway (The gene in the blue frame was predicted major gene, and the gene in the purple frame was the major gene identified by previous study, and the gene behind orange point was transcription factor. The arrow means the possible way that major genes regulate isoflavone).

stable; 2) the effects of these modules on isoflavone are quite stable. Hence, the purple, yellow, turquoise and brown module could regulate the isoflavone accumulation in developing soybean seeds free from the limitation of environments and germplasms.

Furthermore, *via* performing co-expression analysis and transcription factors analysis based on the major genes and molecular modules, four transcription factors were obtained: *Glyma.03G141200* (*bHLH48*), *Glyma.07G054000* (*MYB111*), *Glyma.14G089600* (*bHLH74*) and *Glyma.17G160500* (*MYB85*) (Table S10 and Figure 4A). The following evidences in this study support that these transcription factors regulate the accumulation of isoflavones: 1) significantly differential expression were existed among high isoflavone lines and low isoflavone lines; 2) significantly differential expression were found during seeds development and accompanied by isoflavone variation; 3) GS analysis examined that these genes were highly correlated with the accumulation of various isoflavones; 4) according to the results of transcription factor target genes analysis, many major genes were regulated by them; 5) the strong correlation between transcription factor and target genes could be found in meta-analysis as well. Among them, *MYB111* has been demonstrated to regulate flavonoid synthesis in *Arabidopsis thaliana* by Stracke et al. (2007), while the other 3 genes were found for the first time.

3.4.2 QTG analysis and isoflavones biosynthesis pathway expansion

Comparing the major genes with the 63 stable QTL (52 meta-QTL and 11 novel QTL) mapped in our previous study (Table S2) (Chen et al., 2021), 7 major genes were identified to be the QTG of 10 stable QTL (Figure 4B). Among them, 3 QTG encodes transcription factors, while 3 QTG encodes CYP450 protein. The QTL, qGLEME031, contains 2 QTG, *Glyma.03G141200* (*bHLH48*) and *Glyma.03G122000* (*CYP450 98A2*). *Glyma.07G054000* (*MYB111*) is the QTG of 3 stable QTL.

Additionally, on the basis of their effects and functions, these major genes could be divided into 5 categories: 1) 13 genes could directly take part in the biosynthesis of isoflavones; 2) 7 genes could regulate the biosynthesis of isoflavones directly by transcription regulation or post-transcription regulation; 3) 15 genes could indirectly influence isoflavone accumulation *via* regulating the biosynthesis of other related metabolites (such as anthocyanin); 4) 3 genes could mediately underly isoflavone synthesis through regulating the soybean seed growth and development; 5) the effects of 3 genes are unknown currently. Thus, *via* integrating the functions and internal interaction of major genes, the isoflavones biosynthesis pathway in soybean was supposed to be expanded (Figure 4C).

3.4.3 The expression of major genes in soybean seeds during seed development

A series of qRT-PCR assays showed that the major genes were constitutively expressed in the seeds of the ZH04 (low total isoflavone content) and ZH05 (high total isoflavone content) lines (Figure 5). Obviously, the expression trends of these major genes are consistent with the variation of isoflavone content. This result indicates that the expression of major genes in soybean seeds is related to isoflavone content, which also suggested that our results are reasonable and reliable.

3.4.4 Dynamic molecular mechanisms regulating isoflavone accumulation in developing seeds

Combining with the major genes and accumulation pattern, the process of isoflavone accumulation can be divided into three phases: phase I (S1-S3, the initiation of synthesis), phase II (S4-S6, the main synthesis and rapid variation) and phase III (S7-S9, the termination of synthesis).

With further combination of the association analysis, the dynamic molecular patterns regulating isoflavones accumulation in developing soybean seed are characterized in Figure 6. There are three obvious rules: 1) the genes that have the most impact vary from time to time; 2) a single gene might show antipodal effects in different phase; 3) some genes modulate isoflavones accumulation in one phase, while their homologous genes might regulate the accumulation in another phase. The internal reasons of these rules

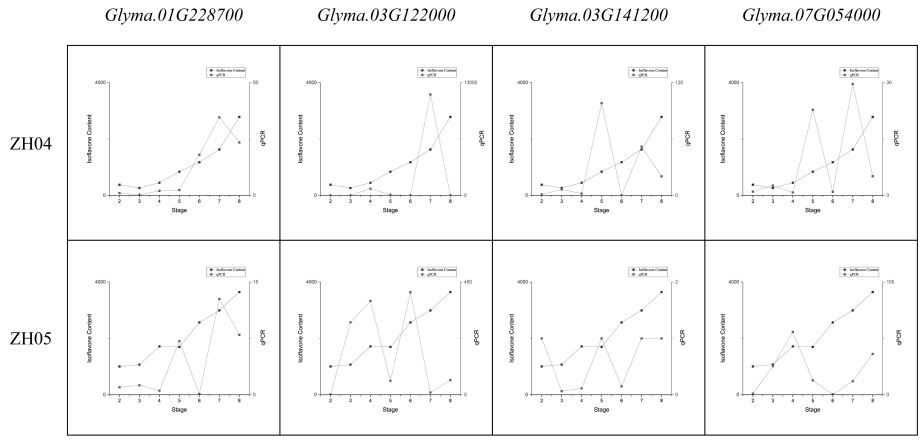


FIGURE 5
Comparing the variation of major genes expression and total isoflavone content during seed development.

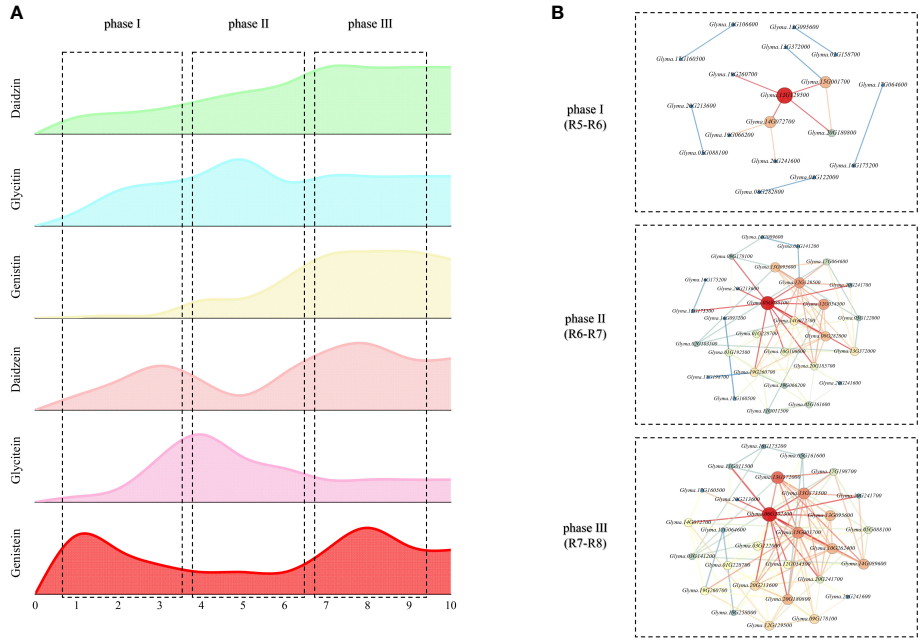


FIGURE 6
Dynamic molecular mechanisms regulating isoflavone accumulation in developing seeds **(A)** Trend of isoflavone dynamic variation; **(B)** Dynamic molecular network regulating isoflavone.

are suggested as follows: 1) the variation of key genes might be depended on the plant growth factors, like phytohormone, for the main life activity of plants is determined by their growth stage whose variation are controlled by plant growth factors; 2) the antipodal effects of same genes might be developed by the negative control mechanisms of plants; 3) the temporal variation of homologous genes may be regulated by photoperiod signal transduction pathway.

In a word, this research: 1) characterized the dynamic variation of isoflavone content; 2) obtained 41 stable major genes and 4 molecular modules influencing isoflavone accumulation; 3) expanded the isoflavone biosynthesis pathway; 4) explored the dynamic molecular patterns regulating isoflavones accumulation in the developing seed of soybean.

4 Discussion

4.1 Dynamic variation of six isoflavone components

During soybean seed development, there are huge and significant differences in the dynamic variation of various isoflavones.

The accumulation pattern of daidzin is “climbing mode”. Three noticeable characteristics of its variation are suggested: 1) the first gap between high daidzin contents lines and low daidzin contents lines generally occurred in S4 period; 2) the higher daidzin contents lines could maintain their biosynthesis rates, while the rate of lower lines decreased at S4; 3) the accumulation rate of high daidzin contents lines was almost equivalent to that of the low daidzin contents lines, but the accumulation duration in higher lines was obviously longer than lower lines. Thus, the uppermost reason determining the daidzin contents in matured seeds of soybean should be the accumulation duration, rather the accumulation rates. With further speculation, the current research suggests: 1) the daidzin contents could be regulated by modulating seed growth period; 2) the expression of daidzin related genes in different lines is similar in early developing stage, but differs in late developing stage.

The accumulation pattern of daidzein is “two peaks mode”. According to the variation of daidzein content in different lines, there is no significant correlation between the peak value and final content. For example, the peak value of HF25 (the parent with lower daidzein content) was noticeably greater than the peak value of ZD27 (the parent with higher daidzein content). Additionally, the germplasms with high daidzein contents generally reached their peak later than the low-daidzein content germplasms did. Consequently, summarizing these phenomena, this study demonstrates that: 1) the daidzein contents in matured seed might be determined by not only the biosynthesis of daidzein, but also the decomposition and metabolism, and latter might be more important; 2) most of daidzein synthesized in the early stage of seed development would be further metabolized, while the daidzein accumulated in late stage of seed development would be stored; 3) daidzein contents are modulated by growth-regulating factors (GRF), as same as daidzin contents are.

The accumulation pattern of glycitin is “single peak and climbing mode”. An obviously characteristic is that the glycitin contents reach equilibrium at S6, siginificantly earlier than other isoflavone components. This indicates that the accumulation of glycitin is a short-time activity and centered in the early development stage. In addition, the glycitin contents of parents were similar, but a significant phenotype variation is found in their

RIL lines. This suggests that: 1) although the parents are similar at phenotype level, their genotypes may be diverse; 2) the variation within group is quite large, and a lot of lines with different genotypes are generated; 3) many glycitin related genes may be mutated.

The accumulation pattern of glycitein is "single peak mode". A strong correlation is existed between seed glycitein contents and length of soybean growth period, which could be inferred from two special samples: ZD27 and ZH04. Generally, glycitein contents increased during S1-S4, and decreased slowly during S4-S7, and then kept constant till the end. However, the variation of glyceitin contents in ZD27 started at S2 and ended at S5, which exhibits a short and early variation process; while ZH04 had a long and late accumulation process which reached the peak at S7 and ended at S8. The glyceitin contents of ZD27 is lower than average, while the glyceitin content of ZH04 is the highest one in this research. We infer that the distinction of growth duration period leads to the diverse accumulation processes of glycitein in different germplasms. Indeed, two rules might be suggested: 1) the glycitein contents are depended on the ratio of synthesis rate and resolving rate; 2) in accordance with the accumulation patterns of daidzein, the glycitein synthesized in early developing stage tends to be further metabolized, while the glycitein accumulated in late developing stage is apt to storage. Since the glycitein contents rapidly decreased during S4-S7, but other isoflavones did not variate in this way, a specific regulation mechanism might regulate this process and does not influence the accumulation process of other isoflavones.

The accumulation pattern of genistin is "two-part climbing mode". According to the variation tendency, the gap of genistin contents in different lines starts to increase at S4, and the genistin contents in matured seed is determined by its total synthesis in the rapid accumulation phase (S5-S7) in a large part. In this period, the lines with higher genistin content produced more genistin isoflavone, while the lower genistin lines would accumulate less genistin content. Additionally, the accumulation of genistin is relatively later than other components, which is a typical metabolite synthesized in the late developing stage. The synthesis rate of this kind of metabolites would be significantly influenced by their precursors' contents accumulated in early developing stage. With the combination of the dynamic variation of genistein this study supports that the biosynthesis of genistin depends on the genistein content accumulated during S1-S4. With further speculation, the expression of genistin related genes would be higher after S3, but lower before S3.

The accumulation pattern of genistin is "two-part climbing mode". As its variation in developing seed shown, three rules could be summed up: 1) lines whose first peak is high usually contain lower genistin contents finally, while lines whose first peak is low always contains a higher genistein contents in the end; 2) the first peak of each line almost occurs at S1, but their second peak distributes at various stages; 3) the second peak value and final contents would be higher if the second peak occurs later.

Although there are significant differences in the accumulation patterns of various isoflavones, several common rules appropriate to all components are found. First of all, the biosynthesis of all isoflavones is mainly took place during S2-S8, and the synthesis of aglycones is lightly earlier than of glucosides generally. Secondly, the first gap of contents between high isoflavone line and low isoflavone line appears during S2-S4. Additionally, most of isoflavone synthesized in early development stage would be further metabolized and/or decomposed, while isoflavone synthesized in late development stage would be stored. Finally, one of the most important factors determining isoflavone contents in matured seed is the duration of synthesis during the soybean development. Generally, high isoflavone cultivars always have long synthesis duration.

4.2 Major genes and molecular modules related to isoflavone contents

4.2.1 Major genes and their effects

By integrating transcriptome analysis and meta-analysis, a total of 41 major genes influencing the accumulation of isoflavones were suggested, including 28 novel major genes.

Four transcription factor genes, *MYB111* (*Glyma.07G054000*), *bHLH48* (*Glyma.03G141200*), *bHLH74* (*Glyma.14G089600*) and *MYB85* (*Glyma.17G160500*), were attained. *MYB111* has been demonstrated to regulate flavonoid synthesis in *Arabidopsis thaliana* (Stracke et al., 2007), but its downstream genes and related mechanisms were still unknown. Based on the TF analysis and co-expression analysis, this study argues that its impacts on isoflavone accumulation might be *via* regulating *4CL1*. The other three TF genes were detected for the first time, and related mechanisms are predicted as: 1) the effects of *bHLH48* might be mediated by *Glyma.10G106600* (beta-glucosidase BoGH3B isoform X1) regulating the contents of UDP-glucosides; 2) *bHLH74* should regulate *CHS7*; 3) *MYB85* should indirectly affect isoflavone accumulation *via* regulating anthocyanidin pathway mediated by *Glyma.18G258000* (encoding malonyl-CoA: anthocyanidin 5-O-glucoside-6'-O-malonyltransferase).

One interesting finding is that lots of peroxidase-related genes are suggested to be major genes in the current study, such as peroxidase 51 (*Glyma.01G192500*). The dynamic variation and expression patterns of these major genes are conformed to the definition of major genes influencing isoflavone synthesis. Thus, we identified the peroxidase-related genes affecting isoflavone accumulation in plants. This study also found that: 1) the up-regulated expression of these genes is always accompanied by the decrease of the contents of isoflavone; 2) based on the GS analysis, the gene is strongly negatively correlated with the accumulation of all isoflavones; 3) these genes are expressed together with *Glyma.19G260700* (Caffeic acid 3-O-methyltransferase isoform B), *Glyma.20G185700* (leucoanthocyanidin reductase) and *Glyma.14G072700* (dihydroflavonol-4-reductase DFR1). Thus, three points of views are put out: 1) the effects of peroxidase-related genes in isoflavone accumulation were negative; 2) these genes should influence isoflavone synthesis *via* regulating the pathway of *p*-cumaric acid synthesizing to Naringenin (including *p*-cumaric acid, *p*-cumaroyl-CoA, Chalcone, Naringenin-chalcone and Naringenin); 3) the negative effects of these genes should be resulted by promoting the

intermediate metabolites of isoflavone pathway (like *p*-cumaric acid, *p*-cumaroyl-CoA, Chalcone, Naringenin-chalcone and Naringenin) synthesizing to other secondary metabolites, such as lignin, flavanol and anthocyanidin.

Another interesting finding is centered by *IFS1* (*Glyma.07G202300*) and *IFS2* (*Glyma.13G173500*). *IFS* is a quite important gene for isoflavone synthesis, for it 1) encodes protein catalyzing the first committed step of isoflavone biosynthesis; 2) determines whether plants produce isoflavone (Jung et al., 2000). Hence, the intuitive view, up-regulated expression of *IFS* should lead to higher production of isoflavone in soybean seed, and could be easy to accepted. In this study, however, neither *IFS1* nor *IFS2* are not suggested to have great impacts in improving the production of isoflavones. We found that they are notably highly expressed in ZD27 during S5-S7 comparing with ZH01, ZH02 and HF25, while the isoflavone content of ZD27 is higher than ZH02 and HF25, but equal to ZH01. Additionally, the highly expressive stages of *IFS* were later than the stages of isoflavones rapid accumulation, and weak correlation between isoflavones and *IFS* are found. With the combination of the above views and phenomenon, two assumptions are proposed: 1) *IFS* does determine isoflavone synthesis in plants, but does not regulate isoflavone contents; 2) another *IFS* homologous gene are existed, whose high expression could offset low expression levels of *IFS1* and *IFS2*. We tend more to support the first hypothesis, because *IFS* homologous gene is not found in this study. Though several major genes detected in this study are function-unknown gene, the results of sequence alignment between *IFS* and them do not support they are homologous. With further deduction on the basis of the first assumption, isoflavone synthase may have a long median life expectancy and could catalyze the biosynthesis of isoflavone rapidly, and therefore the pathway does not require too many IFS proteins. Consequently, the activity and number of IFS would not restrict the rate of isoflavone synthesis in developing soybean seeds generally, and up-regulated expression of *IFS* would not lead to higher production of isoflavone.

4.2.2 Potentially new regulation mechanisms of isoflavone synthesis

The isoflavone synthesis is regulated by multiple factors, here we propose two potential regulation mechanisms below.

The rules of isoflavones dynamic change in this study suggests that a strong correlation should be existed between the seed development and isoflavone accumulation. Several evidences in this study imply that the essence of this relationship is photoperiod regulating isoflavone accumulation, and related mechanism might be mediated by 3 major genes, *CHS4a* (*Glyma.08G110500*), *CHS7* (*Glyma.01G228700*) and *bHLH74* (*Glyma.14G089600*). First of all, compared with the early-matured cultivars and lines (like HF25, SN14 and ZH02), the expression abundance of *CHS* is obviously high in late matured samples (like ZD27 and Wm82). In addition, *CHS* is involved in isoflavone synthesis and circadian clock, according to the gene functional annotation. The expression correlation analysis also suggests that several circadian clock genes are significantly correlatively expressed with *CHS7*, such as *Glyma.07G048500* (*Late-elongated hypocotyl*, *LHY*) and *Glyma.14G049700* (encoding E3 ubiquitin-protein ligase COP1). Additionally, this study also found that *CHS7* and these circadian clock genes were regulated by *bHLH74*. Actually, the view, *CHS* involved in photoperiod, has been identified in *Arabidopsis*. Jenkins et al. (2001), based on the experiments with photoreceptor mutants, discovered that distinct UV-A/blue (cry mediated) and UV-B photoreception systems control *CHS* expression. Hildreth et al. (2022) found that the transcriptional activity of two key circadian clock genes, *CCA1* (*Circadian clock associated 1*, also known as *LHY*) and *TOC1* (*Timing of cab expression 1*), was altered in *CHS*-deficient seedlings across the day/night cycle. Several research indirectly supported that *bHLH74*, a circadian clock related TF, influences isoflavone accumulation were also found. Li et al. (2019) demonstrated that the down-regulation of *bHLH74* would inhibit anthocyanin (a secondary metabolite, which has a common precursor, chalcone, with isoflavones) accumulation in *Arabidopsis*. *Via* putting the above points and evidences together, the possible molecular mechanism of photoperiod regulating isoflavone accumulation is inferred: photoreception systems, *via* activating the downstream signal transduction pathway, up-regulate *bHLH74* to promote the expression *CHS*, and therefore accelerate isoflavone accumulation.

Another possible new mechanism modulating isoflavone synthesis was mediated by *4CL1* and *4CL2*. According to their tissues-gene expression atlas (Figure S11), drawn on the basis of Soybean Expression Atlas database (Machado et al., 2020), we found:1) *Gm4CL1* (*Glyma.15G001700*) and *Gm4CL2* (*Glyma.13G372000*) are tissue-specific genes; 2) the expression of *Gm4CL1* and *Gm4CL2* could be also approximative in some tissues, such as flower, pod and seed. In addition, according to Lindermayr et al. (2002) and Schneider et al. (2003), *Gm4CL1* takes part in the biosynthesis of lignin, while *Gm4CL2* takes part in the biosynthesis of isoflavone in soybean. Hence, three opinions are proposed: 1) the functions of *Gm4CL1* and *Gm4CL2* are different; 2) their expression and effects might not be mutual interference; 3) the impacts of *Gm4CL1* on isoflavones are weak. In this study, however, *Gm4CL1* is not only negatively correlated with isoflavone accumulation in soybean, but also significantly up-regulated expressed in HF25 (low isoflavone contents germplasm) comparing with ZD27 and ZH01 (high isoflavone contents germplasms). While *Gm4CL2* is neither strongly correlated with isoflavones accumulation, nor significantly differently expressed among high and low isoflavone contents germplasms. Moreover, the expression of *Gm4CL2* is similar in ZD27 and HF25, while of *Gm4CL1* is obviously up-regulated in HF25 (low isoflavone content cultivars) at S5 and S7 (two important stages of isoflavone synthesis) comparing with ZD27 (high isoflavone content cultivars). Thus, this study proposes that *Gm4CL1* could strongly and negatively regulate isoflavone accumulation in soybean, and the negative effects of *Gm4CL1* might be generated by 4CL1 protein competing substrates with 4CL2. To be more specific, 4CL1 and 4CL2 synthesize different products *via* bonding same substrates (*p*-cumaric acid and *p*-cumaric-CoA), and more substrates would be utilized by 4CL1 when the expression of *Gm4CL1* is increased, and therefore less precursors of isoflavones would be generated by 4CL2 with decreased isoflavones.

4.3 Summary and further research avenue

Based on the developing seeds of soybean from 2 parents and 6 RILs, and combined with the reported QTL and meta-analysis method, this study: 1) characterized the dynamic changes of isoflavone contents; 2) revealed the accumulation patterns of six isoflavone compounds at both physiological and molecular level; 3) obtained 28 novel major genes and 4 molecular modules associated with isoflavones accumulation; 4) expended the isoflavone biosynthesis pathway. The results will give a meaningful guidance for the molecular design and breeding of elite soybean cultivars with high isoflavone contents. Meanwhile, investigating the functional differences between *Glyma.15G001700* (*4CL1*) and *Glyma.13G372000* (*4CL2*), and the molecular mechanisms of growth-regulating factors (GRF) modulating the isoflavone contents are required for future research.

Data availability statement

The data presented in the study are deposited in the NCBI repository, accession number PRJNA895728.

Author contributions

HC and QYZ designed the research; HC, CKL, XW and YSL conducted the experiments and analyzed the data; FFW and XWP conducted the field trial; HC and QYZ wrote and revised the manuscript. These authors' contribution is equally: HC and CKL. All authors read and approved the manuscript.

Funding

This study was financially supported by Strategic Priority Research Program of Chinese Academy of Sciences (XDA24030403-3), and the National Key R&D Program of China (2021YFD1201103-03), and the Project for high-tech industrialization of science and technology between Jilin province and Chinese Academy of Sciences (2022SYHZ0039).

Acknowledgments

Great thanks for Dr. Yingpeng Han (Northeast Agricultural University) providing materials.

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1014349/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Dynamic variation of reported genes associated with isoflavone accumulation.

SUPPLEMENTARY FIGURE 2
The quality control, evaluation and quantification of RNA-seq data **(A)** The histogram of data filtering; **(B)** The histogram of data comparative analysis; **(C)** The histogram of mapped reads distribution; **(D)** The graph of gene covering analysis; **(E)** The box plot of gene expression distribution.

SUPPLEMENTARY FIGURE 3
Verification of RNA-Seq results by qRT-PCR.

SUPPLEMENTARY FIGURE 4
Distribution heap diagram of DEGs.

SUPPLEMENTARY FIGURE 5
Results of WGCNA in this study **(A)** Gene dendrogram and module colors; **(B)** Correlation between modules; **(C)** Correlation between module and trait; **(D)** Correlation between gene and trait; **(E)** Module significance (GS) between module and trait; **(F)** Metabolite pathways of molecular module.

SUPPLEMENTARY FIGURE 6
Results of time series analysis.

SUPPLEMENTARY FIGURE 7
Transcriptome analysis of G1 meta-data.

SUPPLEMENTARY FIGURE 8
Transcriptome analysis of G2 meta-data.

SUPPLEMENTARY FIGURE 9
Transcriptome analysis of G3 meta-data.

SUPPLEMENTARY FIGURE 10
Transcriptome analysis of G4 meta-data.

SUPPLEMENTARY FIGURE 11
Expression level of *Gm4CL1* and *Gm4CL2* in various tissues.

SUPPLEMENTARY FIGURE 12
GO annotation and enrichment analysis of Yellow molecular module.

SUPPLEMENTARY FIGURE 13
GO annotation and enrichment analysis of Purple molecular module.

SUPPLEMENTARY FIGURE 14
GO annotation and enrichment analysis of Brown molecular module.

SUPPLEMENTARY FIGURE 15
GO annotation and enrichment analysis of Turquoise molecular module.

References

Borenstein, M., Hedges, L. V., Higgins, J. P. T., and Rothstein, H. R. (2009). How a meta-analysis works. *Introduction to Meta-Anal.* 1–7. doi: 10.1002/9780470743386.ch1

Chen, H., Pan, X., Wang, F., Liu, C., Wang, X., Li, Y., et al. (2021). Novel QTL and meta-QTL mapping for major quality traits in soybean. *Front. Plant Sci.* 12, 774270. doi: 10.3389/fpls.2021.774270

D'Agostina, A., Boschin, G., Resta, D., Annicchiarico, P., and Arnoldi, A. (2008). Changes of isoflavones during the growth cycle of lupinus albus. *J. Agric. Food Chem.* 56, 4450–4456. doi: 10.1021/jf8003724

Dhaubhadel, S., Gijzen, M., Moy, P., and Farhangkhoee, 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

Dhaubhadel, S., McGarvey, B. D., Williams, R., and Gijzen, M. (2003). Isoflavonoid biosynthesis and accumulation in developing soybean seeds. *Plant Mol. Biol.* 53, 733–743. doi: 10.1023/B:PLAN.0000023666.30358.ae

Dixon, R. A. (2004). Phytoestrogens. *Annu. Rev. Plant Biol.* 55, 225–261. doi: 10.1146/annurev.arplant.55.031903.141729

Ernst, J., and Bar-Joseph, Z. (2006). STEM: a tool for the analysis of short time series gene expression data. *BMC Bioinf.* 7, 191. doi: 10.1186/1471-2105-7-191

Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., et al. (2011). Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* 40, D1178–D1186. doi: 10.1093/nar/gkr944

Graham, T. L., Graham, M. Y., Subramanian, S., and Yu, O. (2007). RNAi silencing of genes for elicitation or biosynthesis of 5-deoxyisoflavonoids suppresses race-specific resistance and hypersensitive cell death in phytophthora sojae infected tissues. *Plant Physiol.* 144, 728–740. doi: 10.1104/pp.107.097865

Guttikonda, S. K., Trupti, J., Bisht, N. C., Chen, H., An, Y.-Q. C., Pandey, S., et al. (2010). Whole genome co-expression analysis of soybean cytochrome P450 genes identifies nodulation-specific P450 monooxygenases. *BMC Plant Biol.* 10, 243. doi: 10.1186/1471-2229-10-243

Hildreth, S. B., Littleton, E. S., Clark, L. C., Puller, G. C., Kojima, S., and Winkel, B. S. J. (2022). Mutations that alter arabidopsis flavonoid metabolism affect the circadian clock. *Plant J.* 110, 932–945. doi: 10.1111/tpj.15718

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

Jenkins, G. I., Long, J. C., Wade, H. K., Shenton, M. R., and Bibikova, T. N. (2001). UV And blue light signalling: pathways regulating chalcone synthase gene expression in arabidopsis. *New Phytol.* 151, 121–131. doi: 10.1046/j.1469-8137.2001.00151.x

Jung, W., Yu, O., Lau, S.-M. C., O'Keefe, D. P., Odell, J., Fader, G., et al. (2000). Identification and expression of isoflavone synthase, the key enzyme for biosynthesis of isoflavones in legumes. *Nat. Biotechnol.* 18, 208–212. doi: 10.1038/72671

Kovaka, S., Zimin, A. V., Pertea, G. M., Razaghi, R., Salzberg, S. L., and Pertea, M. (2019). Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biol.* 20, 278. doi: 10.1186/s13059-019-1910-1

Li, W.-F., Ning, G.-X., Mao, J., Guo, Z.-G., Zhou, Q., and Chen, B.-H. (2019). Whole-genome DNA methylation patterns and complex associations with gene expression associated with anthocyanin biosynthesis in apple fruit skin. *Planta* 250, 1833–1847. doi: 10.1007/s00425-019-03266-4

Lindermayr, C., Möllers, B., Fliegmann, J., Uhlmann, A., Lottspeich, F., Meimberg, H., et al. (2002). Divergent members of a soybean (Glycine max l.) 4-coumarate: coenzyme A ligase gene family. *Eur. J. Biochem.* 269, 1304–1315. doi: 10.1046/j.1432-1033.2002.02775.x

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550. doi: 10.1186/s13059-014-0550-8

Machado, F. B., Moharana, K. C., Almeida-Silva, F., Gazara, R. K., Pedrosa-Silva, F., Coelho, F. S., et al. (2020). Systematic analysis of 1298 RNA-seq samples and construction of a comprehensive soybean (Glycine max) expression atlas. *Plant J.* 103, 1894–1909. doi: 10.1111/tpj.14850

Nueda, M. J., Tarazona, S., and Conesa, A. (2014). Next maSigPro: Updating maSigPro bioconductor package for RNA-seq time series. *Bioinformatics* 30, 2598–2602. doi: 10.1093/bioinformatics/btu333

Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* 14, 417–419. doi: 10.1038/nmeth.4197

Pfaffl, M. W. (2001). A new mathematical model for relative quantification in real-time RT–PCR. *Nucleic Acids Res.* 29, e45–e45. doi: 10.1093/nar/29.9.e45

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

Schneider, K., Hövel, K., Witzel, K., Hamberger, B., Schomburg, D., Kombrink, E., et al. (2003). The substrate specificity-determining amino acid code of 4-coumarate: CoA ligase. *Proc. Natl. Acad. Sci.* 100, 8601–8606. doi: 10.1073/pnas.1430550100

Stark, R., Grzelak, M., and Hadfield, J. (2019). RNA Sequencing: the teenage years. *Nat. Rev. Genet.* 20, 631–656. doi: 10.1038/s41576-019-0150-2

Stracke, R., Ishihara, H., Huep, G., Barsch, A., Mehrtens, F., Niehaus, K., et al. (2007). Differential regulation of closely related R2R3-MYB transcription factors controls flavonol accumulation in different parts of the arabidopsis thaliana seedling. *Plant J.* 50, 660–677. doi: 10.1111/j.1365-313X.2007.03078.x

Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., et al. (2018). STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 47, D607–D613. doi: 10.1093/nar/gky1131

Tian, F., Yang, D.-C., Meng, Y.-Q., Jin, J., and Gao, G. (2019). PlantRegMap: charting functional regulatory maps in plants. *Nucleic Acids Res.* 48, D1104–D1113. doi: 10.1093/nar/gkz1020

Yu, O., Shi, J., Hession, A. O., Maxwell, C. A., McGonigle, B., and Odell, J. T. (2003). Metabolic engineering to increase isoflavone biosynthesis in soybean seed. *Phytochemistry* 63, 753–763. doi: 10.1016/S0031-9422(03)00345-5

Yu, G., Wang, L. G., Han, Y., and He, Q. Y. (2012). ClusterProfiler: An r package for comparing biological themes among gene clusters. *Omics A J. Integr. Biol.* 16, 284–287. doi: 10.1089/omi.2011.0118

Zhang, B., and Horvath, S. (2005). A general framework for weighted gene Co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* 4. doi: 10.2202/1544-6115.1128

Zhao, W., Langfelder, P., Fuller, T., Dong, J., Li, A., and Hovarth, S. (2010). Weighted gene coexpression network analysis: State of the art. *J. Biopharm. Stat* 20, 281–300. doi: 10.1080/10543400903572753

frontiers | Frontiers in Plant Science

TYPE Review
PUBLISHED 14 April 2023
DOI 10.3389/fpls.2023.1150363

OPEN ACCESS

EDITED BY
Chuanen Zhou,
Shandong University, China

REVIEWED BY
Zhengyi Xu,
Northeast Normal University, China
Shuo Li,
Shandong University, China

*CORRESPONDENCE
Xiaoya Lin
xiaoyalin@gzhu.edu.cn
Lin Zhao
zhaolinneau@126.com

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Functional and Applied Plant Genomics, a section of the journal Frontiers in Plant Science

RECEIVED 24 January 2023
ACCEPTED 24 March 2023
PUBLISHED 14 April 2023

CITATION
Yue L, Pei X, Kong F, Zhao L and Lin X (2023) Divergence of functions and expression patterns of soybean bZIP transcription factors.
Front. Plant Sci. 14:1150363.
doi: 10.3389/fpls.2023.1150363

Divergence of functions and expression patterns of soybean bZIP transcription factors

Lin Yue[1†], Xinxin Pei[2†], Fanjiang Kong[1], Lin Zhao[2*] and Xiaoya Lin[1*]

[1]Guangdong Provincial Key Laboratory of Plant Adaptation and Molecular Design, Guangzhou Key Laboratory of Crop Gene Editing, Innovative Center of Molecular Genetics and Evolution, School of Life Sciences, Guangzhou University, Guangzhou, China, [2]Key Laboratory of Soybean Biology of Ministry of Education China, Northeast Agricultural University, Harbin, China

Soybean (*Glycine max*) is a major protein and oil crop. Soybean basic region/leucine zipper (bZIP) transcription factors are involved in many regulatory pathways, including yield, stress responses, environmental signaling, and carbon-nitrogen balance. Here, we discuss the members of the soybean bZIP family and their classification: 161 members have been identified and clustered into 13 groups. Our review of the transcriptional regulation and functions of soybean bZIP members provides important information for future study of bZIP transcription factors and genetic resources for soybean breeding.

KEYWORDS

soybean, transcription factor, bZIP, function, expression patterns

Introduction

Transcription factors (TFs) can be grouped into different families according to their DNA-binding and multimerization domains. Basic region/leucine zipper (bZIP) TFs are characterized by a conserved bZIP domain composed of two motifs: a basic region responsible for binding to specific DNA sequences, and a leucine zipper motif required for dimerization (Hurst, 1995; Wingender et al., 2001; Jakoby et al., 2002). Plant bZIP TFs function in stress and hormone signaling, organ and tissue differentiation, photomorphogenesis, cell elongation, nitrogen/carbon balance, energy metabolism, flower development, seed development, pathogen defense, and gibberellin biosynthesis (Chern et al., 1996; Albani et al., 1997; Oyama et al., 1997; Fukazawa et al., 2000; Jakoby et al., 2002; Cluis et al., 2004; Corrêa et al., 2008; Weltmeier et al., 2009). The functions of plant bZIP proteins appear to be more complex and broader than those of other TFs (Wei et al., 2010).

Due to their crucial roles in numerous biological processes, bZIP TFs have been studied in many plant species: 78 bZIP genes have been identified in *Arabidopsis thaliana* (Dröge-Laser et al., 2018), 92 in rice (*Oryza sativa*) (Corrêa et al., 2008), 125 in maize (*Zea mays*) (Wei et al., 2012), 64 in cucumber (*Cucumis sativus*) (Yoshida et al., 2015a), and 69 in tomato (*Solanum lycopersicum*) (Li et al., 2015). Jakoby et al. (2002) classified *Arabidopsis* bZIP genes into 10 groups (A, B, C, D, E, F, G, H, I, and S) based on the similarity in the basic region and

other conserved motifs. Dröge-Laser et al. (2018) reported four new members in *Arabidopsis*, AtbZIP76-AtbZIP79, and excluded AtbZIP73 as a pseudogene, and classified these 78 AtbZIPs into 13 groups (designated A-M). AtbZIPs are associated with a plethora of functions; most AtbZIPs in each group display group-specific properties (Jakoby et al., 2002; Dröge-Laser et al., 2018).

Soybean (*Glycine max* [L.] Merr) is an important food and industrial crop. Many bZIP genes have been found in soybean. Liao et al. (2008) identified 131 GmbZIP TFs and classified them into 10 groups (A, B, C, D, E, F, G, H, I, and S). Most GmbZIP proteins cluster with the AtbZIP proteins, whereas several GmbZIP members form a distinct S group. Wang et al. (2015) identified 138 GmbZIPs, and Zhang et al. (2018) identified and classified 160 GmbZIPs into 12 groups (A, B, C, D, E, F, G, H, I, J, K, and S). In soybean, 124 and 122 out of 160 GmbZIPs are involved in drought and flooding responses, respectively (Zhang et al., 2018), however, many GmbZIPs have been implicated in various biological processes besides abiotic stress responses. In this review, we analyze the cladistics and expression profiles of GmbZIP and AtbZIP genes, and focus on the well-studied GmbZIP genes, in order to summarize and predict the functions of GmbZIP TFs, and to provide perspectives for their further identification and use in soybean breeding.

Cladistic analysis of GmbZIP proteins

Soybean and *Arabidopsis* bZIP proteins have been recently updated (Sussmilch et al., 2015; Dröge-Laser et al., 2018; Zhang et al., 2018), we regenerated the cladistic tree, and the names of soybean bZIP proteins were based on Zhang et al. (2018) in this review. We aligned the full-length GmbZIP and AtbZIP amino acid sequences by MAFFT v7.505 with default parameters and then conducted a cladistic analysis using iQtree with model JTT+F+R5 and polishing using iTOL (https://itol.embl.de). The GmbZIPs and AtbZIPs classified into 13 groups, which is the same as Zhang et al. (2018) and Dröge-Laser et al. (2018) (Figure 1).

Expression patterns of GmbZIP genes

We obtained expression values (fragments per kilobase of exon per million mapped fragments (FPKMs)) of GmbZIP and AtbZIP genes in different tissues/organs (seeds, flowers, leaves, shoot apices, stems, and roots) from the websites at Phytozome12 (https://phytozome.jgi.doe.gov/pz/portal.html) (Schmutz et al., 2010), and AtGenExpress Plus-Extended Tissue Series in the Arabidopsis eFP Browser (http://bar.utoronto.ca/efp_arabidopsis/cgi-bin/efpWeb.cgi) with the Developmental Baseline as the parameter and other parameters remaining the default (Schmid et al., 2005) (Figure 1). *Arabidopsis* tissues/organs were selected for consisting with soybean tissues/organs. We removed *GmbZIP88*, *AtbZIP18*, *AtbZIP23*, *AtbZIP36*, *AtbZIP62*, *AtbZIP70*, *AtbZIP75*, and *AtbZIP76* due to lack of expression in all tissues/organs. The expression data of the bZIP genes of two species were row scale normalized respectively after $\log2^{FPKM+1}$ by using TBtools (Chen et al., 2020a), and displayed in the heatmap by iTOL (Figure 1; Table S1).

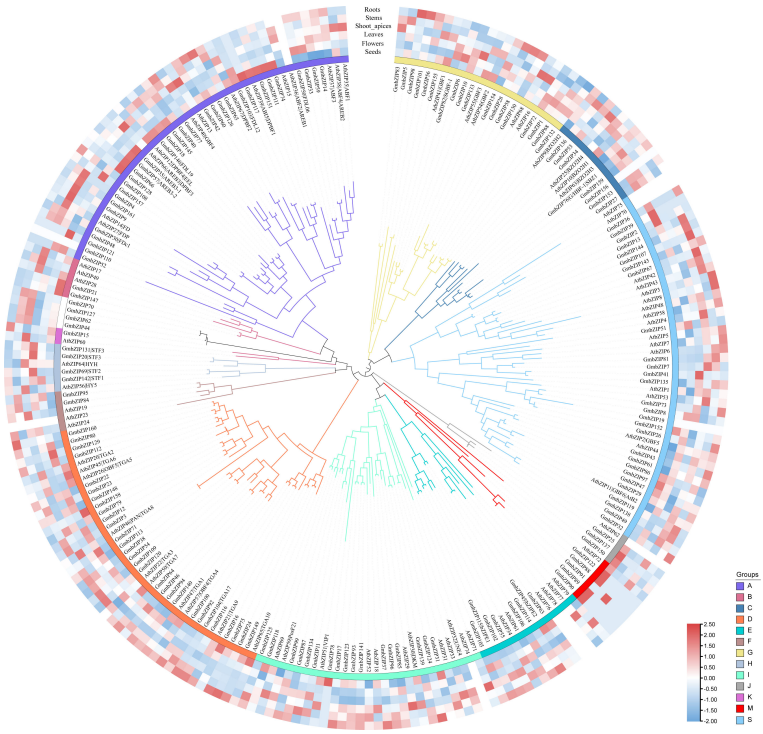


FIGURE 1
Phylogenetic analysis and expression profiles of soybean and Arabidopsis bZIP proteins.

Group A bZIPs regulate abiotic stress responses, plant development, and flowering

We further classified 13 *Arabidopsis* and 31 soybean members in group A, characterized as having conserved motifs containing phosphorylation sites, into four subgroups (Jakoby et al., 2002; Dröge-Laser et al., 2018). The first subgroup contained AtbZIP15 and AtbZIP35-AtbZIP38, which are involved in abscisic acid (ABA) and stress signaling (Choi et al., 2000; Uno et al., 2000). ABA is a plant hormone that regulates diverse processes including stomatal closure, osmotic stress response, and seed maturation and germination (Yoshida et al., 2015b). The members of this first subgroup contained abscisic acid responsive element (ABRE) binding factors (ABFs) that function at the core of ABA signaling (Banerjee and Roychoudhury, 2017). Under osmotic stress conditions (such as drought and high salinity), accumulating ABA is perceived by Pyrabactin resistance 1

(PYR1)/PYR1-LIKE (PYL)/Regulatory components of ABA receptor (RCAR) receptors that inhibit the phosphatase activities of protein phosphatase type 2Cs (PP2Cs). ABA-PYR1/PYL/RCAR-PP2C complexes activate (Sucrose non-fermenting-1) SNF1-related protein kinase 2s (SnRK2s), then SnRK2s directly phosphorylate ABFs to strongly enhance their transactivation properties by directly binding to ABRE *cis*-elements (Yoshida et al., 2015b). We identified four GmbZIP TFs in this subgroup, of which *GmbZIP14*, *GmbZIP59*, *GmbZIP50*, and *GmbZIP33* were mainly expressed in roots and flowers (Figure 1). Because they clustered with AtbZIP35-AtbZIP 38 in the cladistic tree, they likely also participate in ABA and stress signaling. This idea is reinforced by the detailed analysis of GmbZIP14 (name in the reference is listed in Table 1), which is localized in the nucleus and is responsive to ABA, drought, high salinity, and low temperature. Overexpressing *GmbZIP14* improves tolerance to high salt, low temperature, and drought in transgenic plants. Furthermore, some ABREs exist in the promoter region of GmbZIP14 targets, such as *ABA insensitive 1* (*ABI1*), *ABI2*, *RD29B*, *RAB18*, *KAT1*, and *KAT2*, whose expression is also affected by ABA, drought, and high salinity (Gao et al., 2011).

The second subgroup contained ABI5/DPBF1/AtbZIP39, which has been extensively characterized in *Arabidopsis* and functions in ABA-dependent seed maturation and germination (Lopez-Molina et al., 2001; Skubacz et al., 2016). We classified 16 GmbZIP TFs into this subgroup, most of which were highly expressed in seeds (Figure 1), suggesting their potential functions in seed development and germination. In soybean, seed weight is one of the most important yield determinants (Smith and Camper, 1970). The GmbZIP TFs ABA-responsive element binding protein 3-1 (AREB3-1/GmbZIP35), AREB3-2/GmbZIP57, and GmbZIP92 (belonging to Group D) regulate distinct seed development processes by acting with LEAFY COTYLEDON1 (LEC1), a central TF of seed development that controls embryo morphogenesis, photosynthesis, and seed maturation (Jo et al., 2019). LEC1 alone and the LEC1-AREB3 module primarily regulate genes involved in embryo morphogenesis. LEC1-AREB3, LEC1-AREB3-GmbZIP92, and LEC1-AREB3-GmbZIP92-ABI3 modules regulate genes involved in photosynthesis. The LEC1-AREB3-GmbZIP92-ABI3 module also regulates seed maturation genes (Jo et al., 2020) (Figure 2C). Soybean *FD-like 19* (*GmFDL19*, *GmbZIP146*) also classified in this subgroup, is highly induced by ABA, polyethylene glycol (PEG 6000), and high salinity. Overexpressing *GmFDL19* in soybean enhances drought and salt tolerance at the seedling stage. Furthermore, *GmFDL19*/*GmbZIP146* overexpression reduces Na^+ ion accumulation and up-regulates the expression of several ABA- and stress-responsive genes (Li et al., 2017).

The third subgroup, containing FD (AtbZIP14) and FD PARALOG (FDP, AtbZIP27), is involved in control of the *Arabidopsis* floral transition (Abe et al., 2005). The AtbZIP TF FD promotes flowering with the florigen FLOWERING LOCUS T (FT) as a florigen activation complex (Abe et al., 2005; Wigge et al., 2005). TERMINAL FLOWER 1 (TFL1) competes with FT for FD binding and represses the transcription of floral meristem identity genes, such as *LEAFY* (*LFY*) and *APETALA 1* (*AP1*) (Gustafson-Brown et al., 1994; Hanano and Goto, 2011). GmTFL1b (Dt1), the soybean ortholog of *Arabidopsis* TFL1, controls stem growth habit (Liu et al., 2010; Tian et al., 2010) and flowering time (Yue et al., 2021), which strongly influence soybean grain yield (Bernard, 1972; Heatherly and Smith, 2004; Cao et al., 2016). Soybean contains five FD and FDP homologs, which all belong to this subgroup (Sussmilch et al., 2015) (Figure 1). Dt1 interacts with GmFDc1/GmbZIP30 and binds to ACGT *cis*-elements in the promoter region of *GmAP1a* to repress its activity during plant height and flowering time regulation in soybean (Chen et al., 2020b; Yue et al., 2021). Overexpressing *GmFDc1*/*GmbZIP30* leads to early flowering and fewer nodes, suggesting that GmFDc1/GmbZIP30 functions as a floral transition activator (Yue et al., 2021). GmFT5a interferes with Dt1 for GmFDc1 binding and enhances the positive effect of GmFDc1/GmbZIP30 on *GmAP1* expression (Yue et al., 2021). GmFT2a and GmFT5a both induce flowering; however, GmFT5a plays an additional role in termination of shoot apical meristem growth shortly after floral induction (Takeshima et al., 2019). GmFT5a, but not GmFT2a, competes with Dt1 for GmFDc1/GmbZIP30 binding to more rapidly terminate stem growth (Takeshima et al., 2019; Yue et al., 2021) (Figure 2A). The functions of the other four FD and FDP homologs are not known, and can only be inferred from their gene expression patterns: *GmbZIP9* is mainly expressed in shoot apices and seeds, *GmbZIP48*, *GmbZIP30*, *GmbZIP4* and *GmbZIP161* in shoot apices and stems. These distinct expression patterns indicate their functional differentiation during soybean selection and domestication. Furthermore, three other group A members are also involved in floral transition: GmFDL06/GmbZIP50, which interacts with GmFT5a; GmFDL12/GmbZIP103, which functions together with Dt1, GmFT2a, and GmFT5a; and GmFDL19/GmbZIP146, which associates with GmFT2a and GmFT5a (Nan et al., 2014; Takeshima et al., 2019). *GmFDL19*/*GmbZIP146* overexpression in soybean causes early flowering, which may be mediated by upregulation of floral identity genes, such as *Suppressor of overexpression of CO 1* (*SOC1*s), *LFY*s, and *AP1*s, the possible direct targets of GmFDL19/GmbZIP146 (Nan et al., 2014) (Figure 2A).

10.3389/fpls.2023.1150363

TABLE 1 Well studied soybean bZIP transcription factors.

Group ID	Gene name	Gene name in references	Gene ID	Function	Reference
A	GmbZIP14	GmbZIP1	Glyma.02G131700	Salt, drought and low temperature stresses responses	Gao et al. (2011)
	GmbZIP35	AREB3-1	Glyma.04G124200	Seed development	Jo et al. (2020)
	GmbZIP57	AREB3-2	Glyma.06G314400	Seed development	Jo et al. (2020)
	GmbZIP30	GmFDc1	Glyma.04G022100	Flowering; stem growth habit	Yue et al. (2021)
	GmbZIP50	GmFDL06	Glyma.06G040400	Flowering	Takeshima et al. (2019)
	GmbZIP103	GmFDL12	Glyma.12G184432	Flowering	Nan et al. (2014); Takeshima et al. (2019)
	GmbZIP146	GmFDL19	Glyma.19G122800	Flowering; salt and drought stresses responses	Nan et al. (2014); Takeshima et al. (2019); Li et al. (2017)
C	GmbZIP27	GmbZIP105	Glyma.03G247100	Pathogen response	Alves et al. (2015)
	GmbZIP53	GmbZIP62	Glyma.06G079800	ABA signaling; salt and low temperature stresses responses; pathogen response	Liao et al., 2008; Alves et al. (2015)
	GmbZIP76	G/HBF-1, SBZ1	Glyma.10G162100	Pathogen response	Dröge-Laser et al. (1997); Yoshida et al. (2008)
	GmbZIP159	GmbZIP159	Glyma.20G224500	Seed development	Hu et al. (2022)
D	GmbZIP92	GmbZIP67	Glyma.11G183700	Seed development	Jo et al. (2020)
	GmbZIP104	GmTGA17	Glyma.12G184500	Salt and drought stresses responses	Li et al. (2019)
E	GmbZIP45	GmbZIPE2	Glyma.05G168100	Pathogen response	Alves et al. (2015)
G	GmbZIP28	GmbZIP78	Glyma.03G255000	ABA signaling; salt and low temperature stresses responses	Liao et al. (2008)
	GmbZIP82	SGBF-1	Glyma.11G065000	Cold stress response	Kim et al. (2001)
H	GmbZIP142	STF1	Glyma.18G117100	Photomorphogenic; shade avoidance syndrome; lightsignal and nodulation	Shin et al. (2016); Lyu et al. (2021); Ji et al. (2021)
	GmbZIP69	STF2	Glyma.08G302500	Shade avoidance syndrome; light signal and nodulation	Lyu et al. (2021); Ji et al. (2021)
	GmbZIP131	STF3	Glyma.16G092700	Lightsignal and nodulation	Wang et al. (2021)
	GmbZIP20	STF4	Glyma.03G081700	Light signal and nodulation	Wang et al. (2021)
K	GmbZIP15	GmbZIP15	Glyma.02G161100	Salt and drought stresses responses	Zhang et al. (2020)
M	GmbZIP115	GmbZIPE1	Glyma.13G292800	Pathogen response	Alves et al. (2015)
S	GmbZIP8	GmbZIP60	Glyma.02G012700	ABA signaling; salt stress response	Xu et al. (2015)
	GmbZIP32	GmbZIP44	Glyma.04G029600	ABA signaling; salt and low temperature stresses responses	Liao et al. (2008)
	GmbZIP51	GmbZIP2	Glyma.06G048500	Salt and drought stresses responses	Yang et al. (2020)
	GmbZIP61	GmbZIP110	Glyma.08G115300	Salt stress response; root growth	Xu et al. (2016); Manavalan et al. (2015)
	GmbZIP97	GmbZIP97	Glyma.12G040600	Seed development	Hu et al. (2022)
	GmbZIP152	GmbZIP152	Glyma.19G216200	Pathogen response; salt, drought, and heavy metal stress responses	Chai et al. (2022)

The uppercase letters are group IDs.

Group A *Arabidopsis* bZIPs could be divided into four subgroups; three of them have specific functions: ABA and stress signaling, seed maturation and germination, and flowering time and stem growth. Although only a few Group A soybean members are well studied, their functions are largely consistent with those of *Arabidopsis* members. These studies provide directions for studying other members in each subgroup.

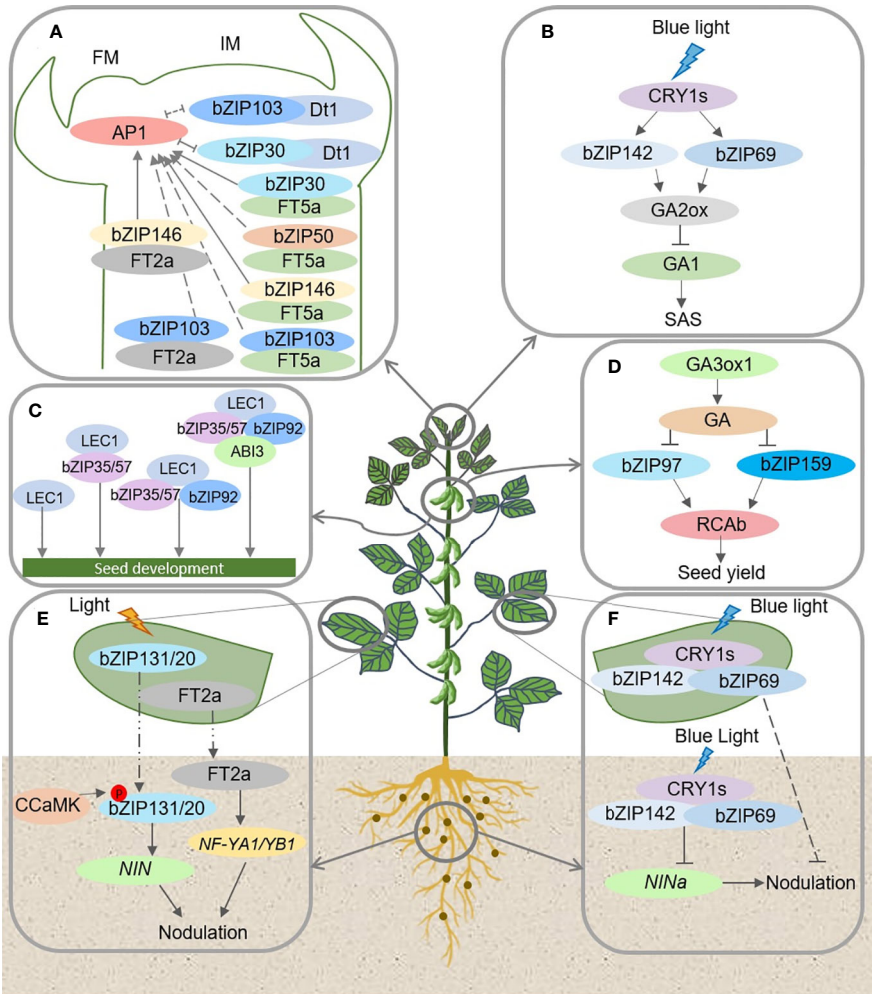


FIGURE 2
Regulation mechanisms of soybean bZIP TFs involved in selected signaling pathways. **(A)** The soybean bZIP TFs FDc1/bZIP30, FDL06/bZIP50, FDL12/bZIP103, and FDL19/bZIP146 function together with Dt1, FT2a, and/or FT5a in the regulation of flowering and stem growth habit. **(B)** STF1/bZIP142 and STF2/bZIP69, whose abundances are increased by light-activated CRY1s, directly upregulate the expression of genes encoding gibberellin 2 oxidases to deactivate gibberellin 1 and control the blue light-induced shade avoidance syndrome. **(C)** The soybean bZIP TFs ABA-responsive element binding protein 3-1 (AREB3-1/bZIP35), AREB3-2/bZIP57, and bZIP92 regulate distinct seed development processes by acting with LEC1. **(D)** bZIP97 and bZIP159 are involved in gibberellin biosynthesis, which is associated with soybean seed yield. **(E)** STF3/bZIP131-STF4/bZIP20-FT module integrates the light-induced shoot signal and the rhizobium-activated root signal, which coordinately promote nodule formation. **(F)** CRY1s interact with and activate STF1/bZIP142 and STF2/bZIP69 transcription in shoots and roots, which repress *NINa* expression, thereby inhibiting nodulation. FM: Floral meristem, IM: Inflorescence meristem, SAS: Shade avoidance syndrome. The solid arrows and the solid blunt ended arrows represent the stimulation and inhibition effect on downstream gene or substances, respectively. The dashed lines represent the stimulation remains to be further confirmed. ⁻·⁻ represent movement of proteins.

Group B and K bZIPs function in endoplasmic reticulum stress pathways

Two bZIPs (AtbZIP17 and AtbZIP28) in group B and the only member in group K (AtbZIP60) function in two endoplasmic reticulum (ER) stress pathways (Iwata and Koizumi, 2005; Howell, 2013). ER stress occurs under adverse environmental conditions, and the ER stress response is implicated in seed development and pathogen response (Vitale and Ceriotti, 2004). AtbZIP17 and AtbZIP28 regulate the unfolded protein response pathway, and AtbZIP60 is involved in a plant-specific ER stress response signal transduction pathway (Iwata and Koizumi, 2005; Howell, 2013). There are two soybean members in group B, GmbZIP21 and GmbZIP147, whose genes are highly expressed in seeds. Their expression patterns provide clues about their analogous functions in the ER stress response; however, their functions are not clearly understood. The only soybean member in group K (GmbZIP15) is involved in the abiotic stress response, different from its *Arabidopsis* homolog AtbZIP60. *GmbZIP15* transcription is suppressed under salt- and drought-stress conditions. Overexpressing *GmbZIP15* in soybean results in hypersensitivity to abiotic stress compared with wild-type plants, which is associated with lower expression of stress-responsive genes, defective stomatal aperture regulation, and reduced antioxidant enzyme activities (Zhang et al., 2020). Considering the crucial functions of AtbZIP60 in the ER stress response, the functions of GmbZIP15 deserve more attention.

Group C and S bZIPs regulate stress responses and plant development

In *Arabidopsis*, four group C members preferentially heterodimerize with five group S1 members, which is defined as the 'C/S1 bZIP network' (Jakoby et al., 2002; Weltmeier et al., 2006). The C/S1 bZIP network is involved in metabolic readjustment during low-energy signaling, downstream of SnRK1 (Baena-González et al., 2007). SnRK1-C/S1 signaling is involved in pathogen defense, which is an energy-consuming process requiring metabolic readjustment. Likewise, C/S1 bZIPs, such as AtbZIP10 in *Arabidopsis* and AtbZIP63 orthologs in several plant species, are implicated in pathogen defense (Dröge-Laser et al., 1997; Kuhlmann et al., 2003; Kaminaka et al., 2006; Shen et al., 2016). We detected 8 and 28 soybean members in groups C and S, respectively, and most of them are expressed in roots and stems, similar to their *Arabidopsis* counterparts (Figure 1). Dröge-Laser et al. (1997) isolated a group C member, G/HBF-1/GmbZIP76, which activates pathogen defense by binding to the promoters of *Chalcone Synthase 15* (*CHS15*), *CHS7*, and *CHS1*, which belong to the chalcone synthase (CHS) family that catalyzes the first and key regulatory step of flavonoid biosynthesis, the well-characterized defense substances (Dixon et al., 1983; Hahlbrock and Scheel, 1989; Wingender et al., 1990; Yoshida et al., 2008). *In-vitro* phosphorylation of G/HBF-1/GmbZIP76 enhances its binding to the *CHS15* promoter. A cytosolic serine kinase that is stimulated by an avirulent strain of the soybean pathogen *Pseudomonas syringae* pv. *glycinea* was identified using recombinant G/HBF-1/GmbZIP76 as a substrate. Stimulation of G/HBF-1/GmbZIP76 kinase activity and G/HBF-1/GmbZIP76 phosphorylation are terminal events in a signal pathway to activate early transcription-dependent plant defense responses (Dröge-Laser et al., 1997). Genes encoding the Group C proteins GmbZIP53 and GmbZIP27 are differentially expressed in the resistant soybean cultivar PI561356 during Asian soybean rust (ASR) infection, which is caused by an obligate biotrophic pathogenic fungus *Phakopsora pachyrhizi* and results in yield losses of up to 80%, indicating their important roles in the response to ASR infection (Patil et al., 1997; Alves et al., 2015).

Moreover, some *Arabidopsis* C and S1 members are also involved in abiotic stress responses, such as AtbZIP1 and its partners AtbZIP53, AtbZIP10, and AtbZIP25 (Sun et al., 2012; Hartmann et al., 2015). In soybean, GmbZIP61, GmbZIP51, GmbZIP8, and GmbZIP32 in groups C and S positively regulate drought and salt stress responses; and GmbZIP53 positively regulates drought, salt, and low-temperature stress responses; GmbZIP152 positively regulate drought, salt, heavy metal, and *Sclerotinia sclerotiorum* stress responses (Liao et al., 2008; Xu et al., 2015; Xu et al., 2016; Yang et al., 2020; Chai et al., 2022). (*GmbZIP61* overexpression in *Arabidopsis* improved salt tolerance by elevating the survival rate, rosette diameter, relative electrolyte leakage, and proline content after a 200 mM NaCl treatment. GmbZIP61 binds to the ACGT motif in promoters and influences the expression of many stress-related genes as well as the accumulation of proline, Na^+, and K^+ (Xu et al., 2016). *GmbZIP51* is induced by multiple abiotic stresses. *GmbZIP51* overexpression in *Arabidopsis* and soybean hairy roots improves tolerance to drought and salt stresses and enhances the expression of the stress-responsive genes *GmMYB48*, *GmWD40*, *Dehydrins 15* (*GmDHN15*), *Glutathione S-transferase 1* (*GmGST1*), and *Late Embryogenesis Abundant* (*GmLEA*) (Yang et al., 2020). GmbZIP8 is implicated in abiotic stress responses. *GmbZIP8* is induced by ABA and salt stress. Promoter analysis indicated that the *GmbZIP8* promoter contains *cis*-acting elements involved in defense and stress responses, such as the ABREs involved in the ABA response and the MYB binding site involved in the drought response. *Arabidopsis* plants heterologously expressing the *GmbZIP8* promoter indicated that *GmbZIP8* is strongly induced by ABA and weakly induced by salt (Xu et al., 2015). Liao et al. (2008) reported that *GmbZIP32* and *GmbZIP53* are differentially regulated by various treatments. *GmbZIP32* is induced by drought, flooding, and salt stress, but moderately induced by ABA treatment; *GmbZIP53* is slightly induced by drought and salt treatments. Transgenic *Arabidopsis* plants overexpressing *GmbZIP32* or *GmbZIP53* had reduced ABA sensitivity, but enhanced salt and low-temperature stress tolerance. *GmbZIP152* is significantly induced by salt, drought, heavy metal, and *S. sclerotiorum* stresses in soybean. Overexpression of *GmbZIP152* in *Arabidopsis* enhances the resistance to the abiotic and *S. sclerotiorum* stresses. ABA-, JA-, ETH-, and SA-induced biotic- and abiotic-related genes, such as *Early Response to Dehydration 1* (*GmERD1*) and *Pathogenesis-related 2* (*GmPR2*), might be the direct targets of GmbZIP152 (Chai et al., 2022).

The C/S1 bZIP network is also implicated in seed development. The Group C members in *Arabidopsis* (AtbZIP10 and AtbZIP25) control seed storage protein biosynthesis; AtbZIP53 activates seed maturation by interacting with AtbZIP10 and AtbZIP25 (Lara et al., 2003; Alonso et al., 2009). GmbZIP97 (Group S) and GmbZIP159 (Group C) are involved in gibberellin biosynthesis, which is associated with soybean hundred-seed weight and seed number. Knocking out *Gibberellin 3-oxidase1* (*GmGA3ox1*), which encodes a key gibberellin biosynthesis enzyme, decreases the content of bioactive gibberellins in leaves while enhancing photosynthesis, thereby promoting seed yield by upregulating *GmbZIP97* (Group S) and *GmbZIP159* (Group C), and then GmbZIP97 and GmbZIP159 directly activate ribulose-1,5-bispho-sphate carboxylase-oxygenase (Rubisco) activases (*GmRCAb*). Further, GmRCAb induces the production of more Rubisco to increase photosynthesis and ensures sufficient energy transport from leaves to seeds (Hu et al., 2022) (Figure 2D).

Because many genes encoding Group S members are expressed in soybean roots, some of them are implicated in regulating root size and architecture, which are important for yield performance (Price et al., 2002; Manavalan et al., 2009). A major quantitative trait locus on chromosome 8 (the Satt315-I locus) controls tap root length, lateral root number, and shoot length in soybean. Eleven TF genes were identified within the confidence interval of this region, among them, the Group S member *GmbZIP61* is highly expressed in the root pericycle and nodules. Pericycle cells form lateral root primordium, which determine lateral root number (Manavalan et al., 2015). *GmbZIP43*, closely related to *GmbZIP61*, is also highly expressed in roots, which suggests its similar function in regulating root architecture.

Arabidopsis Group C members always heterodimerize and work together with Group S1 members; however, this phenomenon has

not been reported in soybean. Therefore, soybean C and S members may have similar biochemical properties, but further study at the molecular level is needed.

Group D members have diverse functions

Characterized by a short zipper domain, a conserved C terminus, and a more variable N terminus, Group D comprises the so-called 'TGACG motif-binding factors' (TGAs), which are further divided into five clades in Arabidopsis (Gatz, 2013). AtTGA1/AtbZIP47 and AtTGA4/AtbZIP57 in Clade I participate in root nitrate uptake, nitrate responses, apoplastic defenses, ER stress responses, salicylic acid biosynthesis, and pathogen defense (Miao et al., 1994; Wang and Fobert, 2013; Alvarez et al., 2014; Sun et al., 2018). The *Arabidopsis* Clade II factors AtTGA2/AtbZIP20, AtTGA5/AtbZIP26, and AtTGA6/AtbZIP45 play crucial roles in systemic acquired resistance and detoxification processes (Xiang et al., 1997; Müller and Berger, 2008; Fu and Dong, 2013). AtTGA3/AtbZIP22 in Clade III is involved in basal pathogen resistance and in mediating phytohormonal cross-talk between salicylic acid and cytokinin (Choi et al., 2010). *Arabidopsis* Clade IV members AtTGA9/AtbZIP21 and AtTGA10/AtbZIP65 regulate anther development (Murmu et al., 2010). The *Arabidopsis* Clade V member AtTGA8/PERIANTHIA (AtPAN/AtbZIP46) controls the formation of floral organ primordia (Chuang et al., 1999; Maier et al., 2011). The soybean members in each clade were expressed in distinct tissues (Figure 1), which suggests their differential roles in soybean. *TGACG motif-binding factor 17* (*GmTGA17*, *GmbZIP104*), encoding a Clade IV protein, is strongly induced by drought and salt stress. Heterologous overexpression of *GmTGA17* in *Arabidopsis* and homologous overexpression in soybean hairy roots enhanced drought and salt tolerance (Li et al., 2019). *GmTGA17/GmbZIP104* is highly expressed in roots, stems, and flowers (Figure 1), which indicated its potential function in another development, similar to AtTGA9/AtbZIP21 and AtTGA10/AtbZIP65.

Group E and M bZIPs regulate pollen wall formation and biotic stress responses

We identified six *Arabidopsis* and four soybean bZIP TFs in Group E. Research on *Arabidopsis* Group E bZIP TFs is limited. Only one study reported that one member from Group E, AtbZIP34, has an essential role in pollen wall formation, and the *atbzip34* mutants show pollen morphology and pollen germination defects. AtbZIP34 is involved in lipid metabolism, cellular transport, and intine biosynthesis by regulating the putative downstream gene *MYB97* (Gibalová et al., 2009). The soybean bZIP TFs GmbZIP106 and GmbZIP114 are closely related to AtbZIP34. *GmbZIP106* and *GmbZIP114* are highly expressed in flowers, suggesting their involvement in floral organ development (Figure 1).

Two other soybean bZIP TF genes, *GmbZIPE1/GmbZIP115* (Group M) and *GmbZIPE2/GmbZIP45* (Group E), are differentially expressed during ASR infection in the resistant soybean cultivar PI561356, indicating that their proteins participate in the response to ASR infection (Alves et al., 2015), but their exact functions and regulation mechanisms need further study. All soybean Group M members are highly expressed in seeds, indicating that they have similar functions (Figure 1).

Group F bZIPs regulate zinc deficiency and salt stress responses

Arabidopsis Group F members AtbZIP19 and AtbZIP23 are essential for adaptation to zinc deficiency in *Arabidopsis* roots (Assuncão et al., 2010). AtbZIP24 is an important regulator of the salt stress response; transcriptional repression of *AtbZIP24* improves salt tolerance in *Arabidopsis* (Yang et al., 2009). GmbZIP84 in this group, whose gene is also highly expressed in roots, may have similar functions as its *Arabidopsis* homologs (AtbZIP19, AtbZIP23, and AtbZIP24).

Group G bZIPs regulate abiotic stress responses

Group G proteins are characterized by a proline-rich N-terminal activation domain (Jakoby et al., 2002). G-BOX-BINDING FACTOR1 (AtGBF1), a well-known *Arabidopsis* Group G bZIP TF, regulates blue-light-dependent hypocotyl expansion, lateral root development, natural senescence, and salicylic acid-dependent pathogen defense (Weisshaar et al., 1991; Schindler et al., 1992; Smykowski et al., 2010; Maurya et al., 2015; Giri et al., 2017). We identified 14 soybean bZIPs in Group G, and Soybean G-box binding factor 1 (SGBF-1, GmbZIP82), GmbZIP6, GmbZIP10, and GmbZIP133 are homologs of *Arabidopsis* GBF1, and their genes are mainly expressed in flowers and leaves (Figure 1). SGBF-1/GmbZIP82 has been extensively studied and participates in abiotic stress responses. SGBF-1/GmbZIP82 binds directly to ABREs in cold-regulated gene promoters. SGBF-1/GmbZIP82 interacts with the C2H2-type zinc finger protein SCOF-1 to up-regulate *COLD-REGULATED* (*AtCOR*) expression and enhance cold tolerance in transgenic *Arabidopsis* (Kim et al., 2001). Another well-studied soybean Group G member is *GmbZIP28*, which is slightly induced by NaCl treatment. Transgenic *Arabidopsis* plants overexpressing *GmbZIP28*showed reduced ABA sensitivity, but increased salt and low-temperature tolerance. GmbZIP28 binds to GLM (GTGAGTCAT), ABRE (CCACGTGG), and PB-like (TGAAAA) *cis*-elements and may function in ABA signaling by upregulating *ABI1* and *ABI2*, and has roles in stress tolerance by regulating various stress-responsive genes (Liao et al., 2008). Also in Group G, GmbZIP154, GmbZIP58, and GmbZIP130 are homologs of AtGBF2 and AtGBF3; and GmbZIP72, GmbZIP1, GmbZIP68, and GmbZIP132 are homologs of AtbZIP16 and AtbZIP68, but their functions are not clearly understood.

Group H bZIPs regulate environmental signaling and carbon-nitrogen balance

Group H contains only two *Arabidopsis* members, ELONGATED HYPOCOTYL5 (AtHY5, AtbZIP56) and HY5 HOMOLOG (AtHYH, AtbZIP64). AtHY5/AtbZIP56 promote photomorphogenesis downstream of phytochromes, cryptochromes, and UV-B photoreceptors, and regulates cell elongation, cell proliferation, chloroplast development, lateral root development, pigment accumulation, and nutrient assimilation (Gangappa and Botto, 2016). AtHYH/AtbZIP64 forms heterodimers with AtHY5/AtbZIP56 and enhances transcriptional activation; AtHYH/AtbZIP64 acts redundantly with AtHY5/AtbZIP56 to regulate hypocotyl growth, lateral root growth, pigment accumulation, and the expression of light-inducible genes (Holm et al., 2002; Gangappa and Botto, 2016). Four soybean bZIP TFs belong to this group. Soybean TGACG-motif-binding factor 1 (STF1, GmbZIP142) and STF2/GmbZIP69, mainly expressed in leaves, are homologs of AtHY5. STF3/GmbZIP131 and STF4/GmbZIP20, mainly expressed in leaves, are homologs of AtHYH (Figure 1).

STF1/GmbZIP142 plays a positive role in photomorphogenesis and phytohormone signaling (Song et al., 2008). The C terminus of STF1/GmbZIP142 complemented the *Athy5 Arabidopsis* mutant phenotype for hypocotyl length, root gravitropic response, and chlorophyll and anthocyanin content, indicating their analogous roles in *Arabidopsis* and soybean. STF1/GmbZIP142 interacts with three B-box zinc finger proteins STO homolog (GmSTH), and GmSTH2 and with Constitutively Photomorphogenic 1 (GmCOP1), which play important roles in light-dependent development and gene expression. The regulatory mechanisms that involve GmCOP1, STF1/GmbZIP142, and the B-box factors in soybean may be similar to those in *Arabidopsis*, including STF1/GmbZIP142, GmSTO, and GmSTH degradation in the dark *via* the GmCOP1-mediated ubiquitination pathway (Shin et al., 2016). In addition, both AtHY5/AtbZIP56 and STF1/GmbZIP142 have strong binding affinity to ACGT-containing elements, suggesting that these two proteins have similar functions and may regulate similar downstream genes (Song et al., 2008).

AtHY5/AtbZIP56 also participates in the shade avoidance response; *AtHY5*/*AtbZIP56* and *AtHYH*/*AtbZIP64* are induced by low red/far-red light ratios (Ciolfi et al., 2013). Soybean displays the classic shade avoidance syndrome, including exaggerated stem elongation, which leads to lodging and yield reduction under dense planting conditions. Two AtHY5/AtbZIP56 homologs in soybean, STF1/*GmbZIP142* and STF2/GmbZIP69, whose abundances are increased by light-activated Cryptochrome Circadian Regulator 1 (GmCRY1s), directly upregulate the expression of genes encoding gibberellin 2 oxidases to deactivate gibberellin 1 and repress stem elongation and control the blue light-induced shade avoidance syndrome. *GmCRY1b* overexpression lines are promising lodging-resistant options for dense planting and intercropping conditions (Lyu et al., 2021) (Figure 2B).

In *Arabidopsis*, another important role of AtHY5/AtbZIP56 is to adjust the carbon-nitrogen balance. In leaves, AtHY5/AtbZIP56 activates the transcription of SWEET-facilitator genes to support sucrose export to roots (Chen et al., 2016). AtHY5/AtbZIP56 moves from shoot to root to activate its own expression to promote nitrate uptake by activating *NITRATE TRANSPORTER2.1* (*AtNRT2.1*) expression (Chen et al., 2016). In roots, AtHY5/AtbZIP56 is involved in nitrogen signaling pathways by positively regulating *NITRATE REDUCTASE2* (*AtNIA2*) and *NITRITE REDUCTASE1* (*AtNIR1*) expression, and negatively regulating *AtNRT1.1* and *AMMONIUM TRANSPORTER1;2* (*AtAMT1;2*) expression (Yanagisawa, 2014; Huang et al., 2015). Different from *Arabidopsis*, legumes evolved a symbiotic relationship with rhizobia, who fix atmospheric nitrogen and provide nitrogen nutrients to their host plant, and the soybean AtHY5 homologs STF1/GmbZIP142 and STF2/GmbZIP69 are also involved in light-mediated symbiotic root nodulation (Wang et al., 2021). Specifically, the blue light receptor GmCRY1-STF1/GmbZIP142-STF2/GmbZIP69 module plays a pivotal role in integrating darkness/blue light and nodulation signals. Soybean perceives blue light by GmCRY1s, which activates *STF1*/*GmbZIP142* and *STF2*/*GmbZIP69* transcription in shoots and roots. Root GmCRY1s interact with and elevate the levels of STF1/GmbZIP142 and STF2/GmbZIP69, which repress *Nodule Inception a* (*GmNINa*) expression, thereby inhibiting nodulation (Ji et al., 2021) (Figure 2F). Wang et al. (2021) demonstrated that light-induced STF3/GmbZIP131, STF4/GmbZIP20, and GmFTs interdependently induce nodule organogenesis from shoots to roots. The rhizobium-activated calcium- and calmodulin-dependent protein kinase (CCaMK) phosphorylates STF3/GmbZIP131, triggering STF3/GmbZIP131-FT2a complex formation, which directly activates *GmNIN* and *Nuclear factor Y* (*GmNF-YA1* and *GmNF-YB1*) expression. The GmCCaMK-STF3/GmbZIP131-STF4/GmbZIP20-FT module integrates the light-induced shoot signal and the rhizobium-activated root signal, which coordinately promote nodule formation (Figure 2E).

Group I bZIPs regulate plant development

We identified 12 *Arabidopsis* and 17 soybean members in Group I, who share a characteristic lysine residue in the basic domain that replaces the highly conserved arginine (Jakoby et al., 2002). The best-studied *Arabidopsis* member in group I is VIRE2-INTERACTING PROTEIN 1 (AtVIP1, AtbZIP51), which is involved in the *Agrobacterium tumefaciens* response, pathogen response, abiotic stress response, cell proliferation, and plant development (Van Leene et al., 2016). The most closely related soybean bZIP TFs are GmbZIP11, GmbZIP134, and GmbZIP78. GmbZIP11 and *GmbZIP134* are highly expressed in leaves and stems; *GmbZIP78* is highly expressed in seeds (Figure 1).

Group I members regulate plant development. AtbZIP29 regulates the cell number in leaves and root meristems by controlling cell wall organization, and DRINK ME (AtDKM, AtbZIP30) affects meristematic tissues and gynoecium development (Lozano-Sotomayor et al., 2016; Van Leene et al., 2016). Their homologous genes in soybean, *GmbZIP85*,

GmbZIP139, and *GmbZIP31*, are highly expressed in flowers, which implies their similar roles as AtDKM, whereas *GmbZIP124* and *GmbZIP96* are highly expressed in roots and stems, which suggests similar functions as those of AtbZIP29. Another Group I member, AtbZIP18, controls pollen development by interacting with AtbZIP34, AtbZIP52, and AtbZIP61, and works redundantly with AtbZIP34 (Gibalová et al., 2017). AtbZIP18, AtbZIP52, AtbZIP34, and AtbZIP61 have one to four homologs in soybean; their unknown redundancy may make it difficult to study their functions. In Group I, AtbZIP59 interacts with lateral organ boundaries domain (LBD) TFs to regulate auxin-induced callus formation, which is required for plant regeneration (Xu et al., 2018). *GmbZIP87*, *GmbZIP99*, *GmbZIP118*, and *GmbZIP125* are mainly expressed in roots and stems, but their exact functions are unknown.

Conclusions and perspectives

Plant bZIP TFs regulate a variety of biological processes. The functions of some soybean bZIP TFs have been extensively studied (Table 1). Most soybean members in Groups B, K, D, E, M, and G regulate abiotic stress responses, which differs from the various functions of *Arabidopsis* members in these groups, such as ER stress response, root nitrate uptake, nitrate responses, apoplastic defenses, pathogen defense, floral organ development, pollen development, lateral root development, and natural senescence. However, the functions of soybean members in Groups A, C, S, and H are highly consistent with those of *Arabidopsis* members in these groups. Our knowledge of soybean members in Groups B, K, D, E, M, and G is limiteds, and reports on the functions of Group F, I, J, and N soybean members are even less clear. Therefore, detailed investigations of all soybean bZIP TF functions and molecular mechanisms are needed. The cladistic, transcriptional, and functional information we provided in each group will be useful for future studies.

Soybean is an important protein and oil crop. Soybean yield is determined by multiple traits, including flowering time, node number, internode length, effective branching number, pod number per plant, seed number per plant, and hundred-seed weight (Pedersen and Lauer, 2004; Lu et al., 2017; Lu et al., 2020). Soybean bZIP TFs regulate many important yield traits. For example, GmFDc1, GmFDL12, GmFDL19, and GmFDL06 (Group A) are transcription cofactors of Dt1, GmFT2a, and/or GmFT5a, which determine stem growth habit and/or flowering time of soybean (Nan et al., 2014; Takeshima et al., 2019; Li et al., 2021; Yue et al., 2021). The detailed functions and regulatory mechanisms of these Group A TFs remain unclear, but they appear to have enormous contributions to soybean yield. STF1 and STF2 (Group H) are involved in the shade avoidance syndrome, which corresponds to internode length of soybean main stems (Lyu et al., 2021). AREB3 (Group A) and GmbZIP92 (Group D) regulate seed development by acting with LEC1 (Jo et al., 2020). GmbZIP97 (Group S) and GmbZIP159 (Group C) are involved in the gibberellin-mediated seed development pathway (Hu et al., 2022). These bZIP TFs are optimal gene resources for soybean breeding. Crop yield is reduced when plants are exposed to extreme environmental conditions such as high salt, drought, cold, and heat, as well as to biotic stresses such as insects and pathogen invasion. Plants exhibit numerous adaptive and protective responses to various abiotic and biotic stimuli. Most of the soybean bZIP TFs are considered abiotic stress regulators; about 75.6% of soybean bZIP genes display transcriptional changes after abiotic stress treatment (Zhang et al., 2018). Among the well-studied soybean bZIP TFs, 11 regulate abiotic stress responses, and 5 are involved in biotic stress responses (Table 1). Most of the environmental-stress-related *Arabidopsis* bZIP TFs are distributed in Groups A, C, S, B, K, D, G, and F. Although the functional diversity and molecular mechanisms of these soybean members need further study, preliminary data provide a valuable basis for future study.

Biological nitrogen fixation is an alternative to nitrogen fertilizer; the ability of legumes to form a symbiosis with nitrogen-fixing rhizobia provides a distinct advantage (Ferguson et al., 2010). STF1, STF2, STF3, and STF4 (Group H) are orthologs of AtHY5 and AtHYH. STF1 and STF2 suppress soybean nodulation, while STF3 and STF4 play positive roles in light-induced nodulation responses in soybean (Ji et al., 2021; Wang et al., 2021). Further studies are needed to determine why these proteins regulate nodulation in different ways, and different strategies should be used to utilize these gene resources in soybean breeding.

The functions of some soybean bZIP TFs have been analyzed *via* physiological experiments and genetic engineering; however, the biochemical properties and regulation mechanisms of many members remain unclear. Because of the duplication of the soybean genome, soybean contains two to ten bZIP TF homologs, which makes it difficult to analyze their homodimers/heterodimers, binding sites, and knockout phenotypes. Advances in genomics and molecular biology have facilitated cloning of homologs and screening of TF binding sites, and clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated protein 9 (Cas9) gene editing technologies have accelerated the generation of single and multiple mutants for studying genes and their functions, which will also facilitate the study of soybean bZIP TFs.

Author contributions

XL, LZ, and FK conceptualized the idea, LY and XP wrote the initial manuscript draft. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the National Natural Science Foundation of China (31901567); the Science and Technology Projects in Guangzhou, China (Grant 202002030180 to XL, Grant 2023A04J1503 to LY); the China Postdoctoral Science Foundation (2019M662843); and the Guangdong Provincial Key Laboratory of Plant Adaptation and Molecular Design.

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1150363/full#supplementary-material

References

Abe, M., Kobayashi, Y., Yamamoto, S., Ichinoki, H., Notaguchi, M., and Goto, K. (2005). FD, a bZIP protein mediating signals from the floral pathway integrator FT at the shoot apex. *Science.* 309, 1052–1057. doi: 10.1126/science.1115983

Albani, D., Hammond-Kosack, M. C., Smith, C., Conlan, S., Colot, V., Holdsworth, M., et al. (1997). The wheat transcriptional activator SPA: A seed-specific bZIP protein that recognizes the GCN4-like motif in the bifactorial endosperm box of prolamin genes. *Plant Cell.* 9, 171–184. doi: 10.1105/tpc.9.2.171

Alonso, R., Oñate-Sánchez, L., Weltmeier, F., Ehlert, A., Diaz, I., Dietrich, K., et al. (2009). A pivotal role of the basic leucine zipper transcription factor bZIP53 in the regulation of arabidopsis seed maturation gene expression based on heterodimerization and protein complex formation. *Plant Cell.* 21, 1747–1761. doi: 10.1105/tpc.108.062968

Alvarez, J. M., Riveras, E., Vidal, E. A., Gras, D. E., Contreras-López, O., Tamayo, K. P., et al. (2014). Systems approach identifies TGA1 and TGA4 transcription factors as important regulatory components of the nitrate response of arabidopsis thaliana roots. *Plant J.* 80, 1–13. doi: 10.1111/tpj.12618

Alves, M. S., Soares, Z. G., Vidigal, P. M. P., Barros, E. G., Poddanosqui, A. M. P., and Aoyagi, L. N. (2015). Differential expression of four soybean bZIP genes during Phakopsora pachyrhizi infection. *Funct. Integr. Genomics* 15, 685–696.

Assuncão, A. G., Herrero, E., Lin, Y. F., Huettel, B., Talukdar, S., Smaczniak, C., et al. (2010). Arabidopsis thaliana transcription factors bZIP19 and bZIP23 regulate the adaptation to zinc deficiency. *Proc. Natl. Acad. Sci. U.S.A.* 107 (22), 10296–10301. doi: 10.1073/pnas.1004788107

Baena-González, E., Rolland, F., Thevelein, J. M., and Sheen, J. (2007). A central integrator of transcription networks in plant stress and energy signalling. *Nature* 448, 938–942. doi: 10.1038/nature06069

Banerjee, A., and Roychoudhury, A. (2017). Abscisic-acid-dependent basic leucine zipper (bZIP) transcription factors in plant abiotic stress. *Protoplasma* 254, 3–16. doi: 10.1007/s00709-015-0920-4

Bernard, R. L. (1972). Two genes affecting stem termination in soybeans. *Crop Sci.* 12, 235–239. doi: 10.2135/cropsci1972.0011183X001200020028x

Cao, D., Takeshima, R., Zhao, C., Liu, B., Jun, A., and Kong, F. (2016). Molecular mechanisms of flowering under long days and stem growth habit in soybean. *J. Exp. Bot.* 8, 1873–1884. doi: 10.1093/jxb/erw394

Chai, M., Fan, R., Huang, Y., Jiang, X., Wai, M. H., Yang, Q., et al. (2022). GmbZIP152, a soybean bZIP transcription factor, confers multiple biotic and abiotic stress responses in plant. *Int. J. Mol. Sci.* 23, 10935. doi: 10.3390/ijms231810935

Chen, C., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y., et al. (2020a). TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant* 13 (8), 1194–1202. doi: 10.1016/j.molp.2020.06.009

Chen, L., Nan, H., Kong, L., Yue, L., Yang, H., Zhao, Q., et al. (2020b). Soybean AP1 homologs control flowering time and plant height. j. integr. *Plant Biol.* 62, 1–12. doi: 10.1111/jipb.12988

Chen, X., Yao, Q., Gao, X., Jiang, C., Harberd, N. P., and Fu, X. (2016). Shoot-to-root mobile transcription factor HY5 coordinates plant carbon and nitrogen acquisition. *Curr. Biol.* 26, 640–646. doi: 10.1016/j.cub.2015.12.066

Chern, M. S., Eiben, H. G., and Bustos, M. M. (1996). The developmentally regulated bZIP factor ROM1 modulates transcription from lectin and storage protein genes in bean embryos. *Plant J.* 10, 135–148. doi: 10.1046/j.1365-313X.1996.10010135.x

Choi, H., Hong, J., Ha, J., Kang, J., and Kim, S. Y. (2000). ABFs, a family of ABA-responsive element binding factors. *J. Biol. Chem.* 275, 1723–1730. doi: 10.1074/jbc.275.3.1723

Choi, J., Huh, S. U., Kojima, M., Sakakibara, H., Paek, K. H., and Hwang, I. (2010). The cytokinin-activated transcription factor ARR2 promotes plant immunity *via* TGA3/NPR1-dependent salicylic acid signaling in arabidopsis. *Dev. Cell.* 19, 284–295. doi: 10.1016/j.devcel.2010.07.011

Chuang, C. F., Running, M. P., Williams, R. W., and Meyerowitz, E. M. (1999). The PERIANTHIA gene encodes a bZIP protein involved in the determination of floral organ number in arabidopsis thaliana. *Genes Dev.* 13, 334–344. doi: 10.1101/gad.13.3.334

Ciolfi, A., Sessa, G., Sassi, M., Possenti, M., Salvucci, S., Carabelli, M., et al. (2013). Dynamics of the shade-avoidance response in arabidopsis. *Plant Physiol.* 163, 331–353. doi: 10.1104/pp.113.221549

Cluis, C. P., Mouchel, C. F., and Hardtke, C. S. (2004). The arabidopsis transcription factor HY5 integrates light and hormone signaling pathways. *Plant J.* 38, 332–347. doi: 10.1111/j.1365-313X.2004.02052.x

Corrêa, L. G. G., Riaño-Pachón, D. M., Schrago, C. G., Vicentini dos Santos, R., Mueller-Roeber, B., and Vincentz, M. (2008). The role of bZIP transcription factors in green plant evolution: adaptive features emerging from four founder genes. *PloS One* 3 (8), e2944. doi: 10.1371/journal.pone.0002944

Dixon, R. A., Dey, P. M., and Lamb, C. J. (1983). Phytoalexins: enzymology and molecular biology. *Adv. Enzymol. Relat. Areas. Mol. Biol.* 55, 1–136. doi: 10.1002/9780470123010.ch1

Dröge-Laser, W., Kaiser, A., Lindsay, W. P., Halkier, B. A., Loake, G. J., Doerner, P., et al. (1997). Rapid stimulation of a soybean protein-serine kinase that phosphorylates a novel bZIP DNA-binding protein, G/HBF-1, during the induction of early transcription-dependent defenses. *EMBO J.* 16, 726–738. doi: 10.1093/emboj/16.4.726

Dröge-Laser, W., Snoek, B. L., Snel, B., and Weiste, C. (2018). The arabidopsis bZIP transcription factor family - an update. *Curr. Opin. Plant Biol.* 45, 36–49. doi: 10.1016/j.pbi.2018.05.001

Ferguson, B. J., Indrasumunar, A., Hayashi, S., Lin, M. H., Lin, Y. H., and Reid, D. E. (2010). Molecular analysis of legume nodule development and autoregulation. *J. Integr. Plant Biol.* 52 (1), 61–76.

Fu, Z. Q., and Dong, X. (2013). Systemic acquired resistance: turning local infection into global defense. *Annu. Rev. Plant Biol.* 64, 839–863. doi: 10.1146/annurev-arplant-042811-105606

Fukazawa, J., Sakai, T., Ishida, S., Yamaguchi, I., Kamiya, Y., and Takahashi, Y. (2000). REPRESSION OF SHOOT GROWTH, a bZIP transcriptional activator, regulates cell elongation by controlling the level of gibberellins. *Plant Cell.* 12, 901–915. doi: 10.1105/tpc.12.6.901

Gangappa, S. N., and Botto, J. F. (2016). The multifaceted roles of HY5 in plant growth and development. *Mol. Plant* 9, 1353–1365. doi: 10.1016/j.molp.2016.07.002

Gao, S. Q., Chen, M., Xu, Z. S., Zhao, C. P., Li, L., Xu, H., 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

Gatz, C. (2013). From pioneers to team players: TGA transcription factors provide a molecular link. *Mol. Plant Microbe Interact.* 26, 151–159. doi: 10.1094/MPMI-04-12-0078-IA

Gibalová, A., Reňák, D., Matczuk, K., Dupl'áková, N., Cháb, D., Twell, D., et al. (2009). AtbZIP34 is required for arabidopsis pollen wall patterning and the control of several metabolic pathways in developing pollen. *Plant Mol. Biol.* 70, 581–601. doi: 10.1007/s11103-009-9493-y

Gibalová, A., Steinbachová, L., Hafidh, S., Bláhová, V., Gadiou, Z., Michailidis, C., et al. (2017). Characterization of pollen-expressed bZIP protein interactions and the role of ATbZIP18 in the male gametophyte. *Plant Reprod.* 30, 1–17. doi: 10.1007/s00497-016-0295-5

Giri, M. K., Singh, N., Banday, Z. Z., Singh, V., Ram, H., Singh, D., et al. (2017). GBF1 differentially regulates CAT2 and PAD4 transcription to promote pathogen defense in arabidopsis thaliana. *Plant J.* 91, 802–815. doi: 10.1111/tpj.13608

Gustafson-Brown, C., Savidge, B., and Yanofsky, M. F. (1994). Regulation of the arabidopsis floral homeotic gene APETALA1. *Cell.* 76, 131–143. doi: 10.1016/0092-8674(94)90178-3

Hahlbrock, K., and Scheel, D. (1989). Physiology and molecular biology of phenylpropanoid metabolism. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 40, 347–369. doi: 10.1146/annurev.pp.40.060189.002023

Hanano, S., and Goto, K. (2011). Arabidopsis TERMINAL FLOWER1 is involved in the regulation of flowering time and inflorescence development through transcriptional repression. *Plant Cell.* 23, 3172–3184. doi: 10.1105/tpc.111.088641

Hartmann, L., Pedrotti, L., Weiste, C., Fekete, A., Schierstaedt, J., Göttler, J., et al. (2015). Crosstalk between two bZIP signaling pathways orchestrates salt-induced metabolic reprogramming in arabidopsis roots. *Plant Cell.* 27, 2244–2260. doi: 10.1105/tpc.15.00163

Heatherly, L. G., and Smith, J. R. (2004). Effect of soybean stem growth habit on height and node number after beginning bloom in the mid-southern USA. *Crop Sci.* 44, 1855–1858. doi: 10.2135/cropsci2004.1855

Holm, M., Ma, L. G., Qu, L. J., and Deng, X. W. (2002). Two interacting bZIP proteins are direct targets of COP1-mediated control of light dependent gene expression in arabidopsis. *Genes Dev.* 16, 1247–1259. doi: 10.1101/gad.969702

Howell, S. H. (2013). Endoplasmic reticulum stress responses in plants. *Annu. Rev. Plant Biol.* 64, 477–499. doi: 10.1146/annurev-arplant-050312-120053

Hu, D., Li, X., Yang, Z., Liu, S., Hao, D., Chao, M., et al. (2022). Downregulation of a gibberellin 3b-hydroxylase enhances photosynthesis and increases seed yield in soybean. *New Phytol.* 235, 502–517. doi: 10.1111/nph.18153

Huang, L., Zhang, H., Zhang, H., Deng, X. W., and Wei, N. (2015). HY5 regulates nitrite reductase 1 (NIR1) and ammonium transporter1;2 (AMT1;2) in arabidopsis seedlings. *Plant Sci.* 238, 330–339. doi: 10.1016/j.plantsci.2015.05.004

Hurst, H. C. (1995). Transcription factors 1: bZIP proteins. *Protein Profile* 2, 101–168.

Iwata, Y., and Koizumi, N. (2005). An arabidopsis transcription factor, AtbZIP60, regulates the endoplasmic reticulum stress response in a manner unique to plants. *Proc. Natl. Acad. Sci. U.S.A.* 102, 5280–5285. doi: 10.1073/pnas.0408941102

Jakoby, M., Weisshaar, B., Droge Laser, W., Vicente Carbajosa, J., Tiedemann, J., Kroj, T., et al. (2002). bZIP transcription factors in arabidopsis. *Trends Plant Sci.* 7, 106–111. doi: 10.1016/S1360-1385(01)02223-3

Ji, H., Xiao, R., Lyu, X., Chen, J., Zhang, X., Wang, Z., et al. (2021). Differential light-dependent regulation of soybean nodulation by papilionoid-specific HY5 homologs, curr. *Biol.* 32, 1–13.

Jo, L., Pelletier, J. M., and Harada, J. J. (2019). Central role of the LEAFY COTYLEDON1 transcription factor in seed development. *J. Integr. Plant Biol.* 61, 564–580. doi: 10.1111/jipb.12806

Jo, L., Pelletier, J. M., Hsu, S. W., Badena, R., Goldberg, R. B., and Harada, J. J. (2020). Combinatorial interactions of the LEC1 transcription factor specify diverse developmental programs during soybean seed development. *Proc. Natl. Acad. Sci. U.S.A.* 117 (2), 1223–1232. doi: 10.1073/pnas.1918441117

Kaminaka, H., Nake, C., Epple, P., Dittgen, J., Schutze, K., Chaban, C., et al. (2006). bZIP10-LSD1 antagonism modulates basal defense and cell death in arabidopsis following infection. *EMBO J.* 25, 4400–4411. doi: 10.1038/sj.emboj.7601312

Kim, J. C., Lee, S. H., Cheong, Y. H., Cheol-Min Yoo, C. M., Lee, S. I., Chun, H. J., et al. (2001). A novel cold-inducible zinc finger protein from soybean, SCOF-1, enhances cold tolerance in transgenic plants. *Plant J.* 25 (3), 247–259. doi: 10.1046/j.1365-313x.2001.00947.x

Kuhlmann, M., Horvay, K., Strathmann, A., Heinekamp, T., Fischer, U., Böttner, S., et al. (2003). The a-helical D1 domain of the tobacco bZIP transcription factor BZI-1 interacts with the ankyrin-repeat protein ANK1 and is important for BZI-1 function, both in auxin signaling and pathogen response. *J. Biol. Chem.* 278, 8786–8794. doi: 10.1074/jbc.M210292200

Lara, P., Onate-Sanchez, L., Abraham, Z., Ferrandiz, C., Diaz, I., Carbonero, P., et al. (2003). Synergistic activation of seed storage protein gene expression in arabidopsis by ABI3 and two bZIPs related to OPAQUE2. *J. Biol. Chem.* 278, 21003–21011. doi: 10.1074/jbc.M210538200

Li, B., Liu, Y., Cui, X. Y., Fu, J. D., Zhou, Y. B., and Zheng, W. J. (2019). Genome-wide characterization and expression analysis of soybean tga transcription factors identified a novel tga gene involved in drought and salt tolerance. *Front. Plant Sci.* 10, 549.

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 (6), e0179554. doi: 10.1371/journal.pone.0179554

Li, X., Fang, C., Yang, Y., Lv, T., Su, T., Chen, L., et al. (2021). Overcoming the genetic compensation response of soybean florigens to improve adaptation and yield at low latitudes, curr. *Biol.* 31, 1–13. doi: 10.1016/j.cub.2021.12.041

Li, D., Fu, F., Zhang, H., and Song, F. (2015). Genome-wide systematic characterization of the bZIP transcriptional factor family in tomato (Solanum lycopersicum l.). *BMC Genom.* 16, 771. doi: 10.1186/s12864-015-1990-6

Liao, Y., Zou, H. F., Wei, W., Hao, Y. J., Tian, A. G., Huang, J., et al. (2008). Soybean GmbZIP44, GmbZIP62 and GmbZIP78 genes function as negative regulator of ABA signaling and confer salt and freezing tolerance in transgenic arabidopsis. *Planta.* 228, 225–240. doi: 10.1007/s00425-008-0731-3

Liu, B., Watanabe, S., Uchiyama, T., Kong, F., Kanazawa, A., Xia, Z., et al. (2010). The soybean stem growth habit gene Dt1 is an ortholog of arabidopsis TERMINAL FLOWER1. *Plant Physiol.* 153, 198–210. doi: 10.1104/pp.109.150607

Lopez-Molina, L., Mongrand, S., and Chua, N. H. (2001). A postgermination developmental arrest checkpoint is mediated by abscisic acid and requires the ABI5 transcription factor in arabidopsis. *Proc. Natl. Acad. Sci. U.S.A.* 98, 4782–4787. doi: 10.1073/pnas.081594298

Lozano-Sotomayor, P., Chávez Montes, R. A., Silvestre-Vañó, M., Herrera-Ubaldo, H., Greco, R., Pablo-Villa, J., et al. (2016). Altered expression of the bZIP transcription factor DRINK ME affects growth and reproductive development in arabidopsis thaliana. *Plant J.* 88, 437–451. doi: 10.1111/tpj.13264

Lu, S. J., Dong, L. D., Fang, C., Liu, S. L., Kong, L. P., Cheng, Q., et al. (2020). Stepwise selection on homeologous PRR genes controlling flowering and maturity during soybean domestication. *Nat. Genet.* 52, 428–436. doi: 10.1038/s41588-020-0604-7

Lu, S. J., Zhao, X. H., Hu, Y. L., Liu, S. L., Nan, H. Y., Li, X. M., et al. (2017). Natural variation at the soybean J locus improves adaptation to the tropics and enhances yield. *Nat. Genet.* 49, 773–779. doi: 10.1038/ng.3819

Lyu, X., Cheng, Q., Qin, C., Li, Y., Xu, X., Ji, R., et al. (2021). GmCRY1s modulate gibberellin metabolism to regulate soybean shade avoidance in response to reduced blue light. *Mol. Plant* 14, 298–314. doi: 10.1016/j.molp.2020.11.016

Maier, A. T., Stehling-sun, S., Offenburger, S., and Lohmann, J. U. (2011). The bZIP transcription factor PERIANTHIA: a multifunctional hub for meristem control. *Front. Plant Sci.* 2, 1–17. doi: 10.3389/fpls.2011.00079

Manavalan, L. P., Guttikonda, S. K., Tran, L. P., 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

Manavalan, L. P., Prince, S. J., Musket, T. A., Chaky, J., Deshmukh, R., Vuong, T. D., et al. (2015). Identification of novel QTL governing root architectural traits in an interspecific soybean population. *PloS One* 10 (3), e0120490. doi: 10.1371/journal.pone.0120490

Maurya, J. P., Sethi, V., Gangappa, S. N., Gupta, N., and Chattopadhyay, S. (2015). Interaction of MYC2 and GBF1 results in functional antagonism in blue light-mediated arabidopsis seedling development. *Plant J.* 83, 439–450. doi: 10.1111/tpj.12899

Miao, Z. H., Liu, X., and Lam, E. (1994). TGA3 is a distinct member of the TGA family of bZIP transcription factors in arabidopsis thaliana. *Plant Mol. Biol.* 25, 1–11. doi: 10.1007/BF00024193

Müller, M. J., and Berger, S. (2008). General detoxification and stress responses are mediated by oxidized lipids through TGA transcription factors in arabidopsis. *Plant Cell.* 20, 768–785. doi: 10.1105/tpc.107.054809

Murmu, J., Bush, M. J., DeLong, C., Li, S., Xu, M., Khan, M., et al. (2010). Arabidopsis basic leucine-zipper transcription factors TGA9 and TGA10 interact with floral glutaredoxins ROXY1 and ROXY2 and are redundantly required for anther development. *Plant Physiol.* 154, 1492–1504. doi: 10.1104/pp.110.159111

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

Oyama, T., Shimura, Y., and Okada, K. (1997). The arabidopsis HY5 gene encodes a bZIP protein that regulates stimulus-induced development of root and hypocotyl. *Genes Dev.* 11, 2983–2995. doi: 10.1101/gad.11.22.2983

Patil, V. S., Wuike, R. V., Thakare, C. S., and Chirame, B. B. (1997). Viability of uredospores of phakopsora pachyrhizi syd. at different storage conditions. *J. Maharashtra Agric. Univ.* 22, 260–261.

Pedersen, P., and Lauer, J. G. (2004). Response of soybean yield components to management system and planting date. *Agron. J.* 96, 1372–1381. doi: 10.2134/agronj2004.1372

Price, A. H., Townend, J., Jones, M. P., Audebert, A., and Courtois, B. (2002). Mapping QTLs associated with drought avoidance in upland rice grown in the Philippines and West Africa. *Plant Mol. Biol.* 48, 683–695. doi: 10.1023/A:1014805625790

Schindler, U., Menkens, A. E., Beckmann, H., Ecker, J. R., and Cashmore, A. R. (1992). Heterodimerization between light-regulated and ubiquitously expressed arabidopsis GBF bZIP proteins. *EMBO J.* 11, 1261–1273. doi: 10.1002/j.1460-2075.1992.tb05170.x

Schmid, M., Davison, T. S., Henz, S. R., Pape, U. J., Demar, M., Vingron, M., et al. (2005). A gene expression map of arabidopsis thaliana development. *Nat. Genet.* 37 (5), 501–506. doi: 10.1038/ng1543

Schmutz, J., Cannon, S., Schlueter, J., Ma, J., Mitros, T., Nelson, W., et al. (2010). Genome sequence of the palaeopolyploid soybean. *Nature* 463, 178–183. doi: 10.1038/nature08670

Shen, L., Liu, Z., Yang, S., Yang, T., Liang, J., Wen, J., et al. (2016). Pepper CabZIP63 acts as a positive regulator during ralstonia solanacearum or high temperature-high humidity challenge in a positive feedback loop with CaWRKY40. *J. Exp. Bot.* 67, 2439–2451. doi: 10.1093/jxb/erw069

Shin, S. Y., Kim, S. H., Kim, H. J., Jeon, S. J., Sim, S. A., Ryu, G. R., et al. (2016). Isolation of three b-box zinc finger proteins that interact with STF1 and COP1 defines a HY5/COP1 interaction network involved in light control of development in soybean. *BBRC.* 478, 1080e1086. doi: 10.1016/j.bbrc.2016.08.069

Skubacz, A., Daszkowska-Golec, A., and Szarejko, I. (2016). The role and regulation of ABI5 (ABA-insensitive 5) in plant development, abiotic stress responses and phytohormone crosstalk. *Front. Plant Sci.* 7, 1884. doi: 10.3389/fpls.2016.01884

Smith, T. J., and Camper, H. M. (1970). Effects of seed size on soybean performance. *Agro. J.* 67, 681–684. doi: 10.2134/agronj1975.00021962006700050025x

Smykowski, A., Zimmermann, P., and Zentgraf, U. (2010). G-Box binding factor1 reduces CATALASE2 expression and regulates the onset of leaf senescence in arabidopsis. *Plant Physiol.* 153, 1321–1331. doi: 10.1104/pp.110.157180

Song, Y. H., Yoo, C. M., Hong, A. P., Kim, S. H., Jeong, H. J., Kim, H. J., et al. (2008). DNA-Binding study identifies c-box and hybrid C/G-box or C/A-box motifs as high-affinity binding sites for STF1 and LONG HYPOCOTYL5 proteins. *Plant Physiol.* 146, 1862e1877. doi: 10.1104/pp.107.113217

Sun, T., Busta, L., Zhang, Q., Ding, P., Jetter, R., and Zhang, Y. (2018). T: GACG-binding factor 1 (TGA1) and TGA4 regulate salicylic acid and pipecolic acid biosynthesis by modulating the expression of systemic acquired resistance deficient 1 (SARD1) and calmodulin-binding protein 60 g (CBP60 g). *New Phytol.* 217, 344–354. doi: 10.1111/nph.14780

Sun, X., Li, Y., Cai, H., Bai, X., Ji, W., Ding, X., et al. (2012). The arabidopsis AtbZIP1 transcription factor is a positive regulator of plant tolerance to salt, osmotic and drought stresses. *J. Plant Res.* 125, 429–438. doi: 10.1007/s10265-011-0448-4

Sussmilch, F. C., Berbel, A., Hecht, V., Vander Schoor, J. K., Ferrándiz, C., Madueño, F., et al. (2015). Pea VEGETATIVE2 is an FD homolog that is essential for flowering and compound inflorescence development. *Plant Cell.* 27 (4), 1046–1046. doi: 10.1105/tpc.115.136150

Takeshima, R., Nan, H., Harigai, K., Dong, L., Zhu, J., Lu, S., et al. (2019). Functional divergence between soybean FLOWERING LOCUS T orthologues FT2a and FT5a in post-flowering stem growth. *J. Exp. Bot.* 70, 3941–3953. doi: 10.1093/jxb/erz199

Tian, Z., Wang, X., Lee, R., Li, Y., Specht, J. E., Nelson, R. L., et al. (2010). Artificial selection for determinate growth habit in soybean. *P. Natl. Acad. Sci. U.S.A.* 107, 8563–8568. doi: 10.1073/pnas.1000088107

Uno, Y., Furihata, T., Abe, H., Yoshida, R., Shinozaki, K., and Yamaguchi-Shinozake, K. (2000). Arabidopsis basic leucine zipper transcription factors involved in an abscisic acid-dependent signal transduction pathway under drought and high-salinity conditions. *P. Natl. Acad. Sci. U.S.A.* 97 (21), 11632–11637. doi: 10.1073/pnas.190309197

Van Leene, J., Blomme, J., Kulkarni, S. R., Cannoot, B., De Winne, N., Eeckhout, D., et al. (2016). Functional characterization of the arabidopsis transcription factor bZIP29 reveals its role in leaf and root development. *J. Exp. Bot.* 67, 5825–5840. doi: 10.1093/jxb/erw347

Vitale, A., and Ceriotti, A. (2004). Protein quality control mechanisms and protein storage in the endoplasmic reticulum. a conflict of interests? *Plant Physiol.* 136, 3420–3426.

Wang, Z., Cheng, K., Wan, L., Yan, L., Jiang, H., Liu, S., et al. (2015). Genome-wide analysis of the basic leucine zipper (bZIP) transcription factor gene family in six legume genomes. *BMC Genom.* 16, 1053. doi: 10.1186/s12864-015-2258-x

Wang, L., and Fobert, P. R. (2013). Arabidopsis clade I TGA factors regulate apoplastic defences against the bacterial pathogen pseudomonas syringae through endoplasmic reticulum- based processes. *PloS One* 8, e77378. doi: 10.1371/journal.pone.0077378

Wang, T., Guo, J., Peng, Y., Lyu, X., Liu, B., Sun, S., et al. (2021). Light-induced mobile factors from shoots regulate rhizobium-triggered soybean root nodulation. *Science.* 374, 65–71. doi: 10.1126/science.abh2890

Wei, K., Chen, J., Wang, Y., Chen, Y., Chen, S., Lin, Y., et al. (2012). Genome-wide analysis of bZIP-encoding genes in maize. *DNA Res.* 19, 463–476. doi: 10.1093/dnares/dss026

Wei, L. Q., Xu, W. Y., Deng, Z. Y., Su, Z., Xue, Y., and Wang, T. (2010). Genome-scale analysis and comparison of gene expression profiles in developing and germinated pollen in oryza sativa. *BMC Genom.* 11, 338. doi: 10.1186/1471-2164-11-338

Weisshaar, B., Armstrong, G. A., Block, A., da Costa e Silva, O., and Hahlbrock, K. (1991). Light-inducible and constitutively expressed DNA-binding proteins recognizing a plant promoter element with functional relevance in light responsiveness. *EMBO J.* 10, 1777–1786. doi: 10.1002/j.1460-2075.1991.tb07702.x

Weltmeier, F., Ehlert, A., Mayer, C. S., Kietrich, K., Wang, X., Schütze, K., et al. (2006). Combinatorial control of arabidopsis proline dehydrogenase transcription by specific heterodimerisation of bZIP transcription factors. *EMBO J.* 25, 3133–3143. doi: 10.1038/sj.emboj.7601206

Weltmeier, F., Rahmani, F., Ehlert, A., Dietrich, K., Schutze, K., Wang, X., et al. (2009). Expression patterns within the arabidopsis C/S1 bZIP transcription factor network: availability of heterodimerization partners controls gene expression during stress response and development. *Plant Mol. Biol.* 69, 107–119. doi: 10.1007/s11103-008-9410-9

Wigge, P. A., Kim, M. C., Jaeger, K. E., Busch, W., Schmid, M., Lohmann, J. U., et al. (2005). Integration of spatial and temporal information during floral induction in arabidopsis. *Science.* 309, 1056–1059. doi: 10.1126/science.1114358

Wingender, E., Chen, X., Fricke, E., Geffers, R., Hehl, R., Liebich, I., et al. (2001). The TRANSFAC system on gene expression regulation. *Nucleic Acids Res.* 29, 281–283. doi: 10.1093/nar/29.1.281

Wingender, R., Rohrig, H., Horicke, C., and Schell, J. (1990). Cis-regulatory elements involved in ultraviolet light regulation and plant defense. *Plant Cell.* 2, 1019–1026.

Xiang, C., Miao, Z., and Lam, E. (1997). DNA-Binding properties, genomic organization and expression pattern of TGA6, a new member of the TGA family of bZIP transcription factors in arabidopsis thaliana. *Plant Mol. Biol.* 34, 403–415. doi: 10.1023/A:1005873500238

Xu, Z., Ali, Z., Xu, L., He, X., Huang, H., Yi, J., et al. (2016). The nuclear protein GmbZIP110 has transcription activation activity and plays important roles in the response to salinity stress in soybean. *Sci. Rep.* 6, 20366. doi: 10.1038/srep20366

Xu, C., Cao, H., Zhang, Q., Wang, H., Xin, W., Xu, E., et al. (2018). Control of auxin-induced callus formation by bZIP59-LBD complex in arabidopsis regeneration. *Nat. Plants.* 4, 108–115. doi: 10.1038/s41477-017-0095-4

Xu, L., Xu, Z., Liu, X., Huang, Y., He, X., Ma, H., et al. (2015). The subcellular localization and ectopic expression analysis in arabidopsis of soybean GmbZIP60 gene. *J. Plant Biochem. Biotechnol.* 24 (1), 9–17. doi: 10.1007/s13562-013-0228-4

Yanagisawa, S. (2014). Transcription factors involved in controlling the expression of nitrate reductase genes in higher plants. *Plant Sci.* 229, 167–171. doi: 10.1016/j.plantsci.2014.09.006

Yang, O., Popova, O. V., Süthoff, U., Lüking, I., Dietz, K. J., and Golldack, D. (2009). The arabidopsis basic leucine zipper transcription factor AtbZIP24 regulates complex transcriptional networks involved in abiotic stress resistance. *Gene.* 436, 45–55. doi: 10.1016/j.gene.2009.02.010

Yang, Y., Yu, T. F., Ma, J., Chen, J., Zhou, Y. B., Chen, M., et al. (2020). The soybean bZIP transcription factor gene GmbZIP2 confers drought and salt resistances in transgenic plants. *Int. J. Mol. Sci.* 21, 670. doi: 10.3390/ijms21020670

Yoshida, T., Fujita, Y., Maruyama, K., Mogami, J., Todaka, D., Shinozaki, K., et al. (2015a). Four arabidopsis AREB/ABF transcription factors function predominantly in gene expression downstream of SnRK2 kinases in abscisic acid signalling in response to osmotic stress. *Plant Cell Environ.* 38 (1), 35–49. doi: 10.1111/pce.12351

Yoshida, T., Mogami, J., and Yamaguchi-Shinozaki, K. (2015b). Omics approaches toward defining the comprehensive abscisic acid signaling network in plants. *Plant Cell Physiol.* 56, 1043–1052. doi: 10.1093/pcp/pcv060

Yoshida, K., Wakamatsu, S., and Sakuta, M. (2008). Characterization of SBZ1, a soybean bZIP protein that binds to the chalcone synthase gene promoter. *Plant Biotechnol.* 25, 131–140. doi: 10.5511/plantbiotechnology.25.131

Yue, L., Li, X., Fang, C., Chen, L., Yang, H., Yang, J., et al. (2021). FT5a interferes with the Dt1-AP1 feed-back loop to control flowering time and shoot determinacy in soybean. *J. Integr. Plant Biol.* 63, 1004–1020. doi: 10.1111/jipb.13070

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, M., Liu, Y., Shi, H., Guo, M., Chai, M., He, Q., et al. (2018). Evolutionary and expression analyses of soybean basic leucine zipper transcription factor family. *BMC Genom.* 19, 159. doi: 10.1186/s12864-018-4511-6

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 17 April 2023
DOI 10.3389/fpls.2023.1163219

OPEN ACCESS

EDITED BY
Zhengjun Xia,
Chinese Academy of Sciences (CAS), China

REVIEWED BY
Ya-nan Jin,
Inner Mongolia Minzu University, China
Yiqiong Huo,
Shanxi Agricultural University, China

*CORRESPONDENCE
Jidao Du
djdbynd@163.com
Qiang Zhao
zqiang0416@hotmail.com

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Functional and Applied Plant Genomics, a section of the journal Frontiers in Plant Science

RECEIVED 10 February 2023
ACCEPTED 30 March 2023
PUBLISHED 17 April 2023

CITATION
Du Y, Zhang Z, Gu Y, Li W, Wang W, Yuan X, Zhang Y, Yuan M, Du J and Zhao Q (2023) Genome-wide identification of the soybean cytokinin oxidase/dehydrogenase gene family and its diverse roles in response to multiple abiotic stress. *Front. Plant Sci.* 14:1163219. doi: 10.3389/fpls.2023.1163219

Genome-wide identification of the soybean cytokinin oxidase/dehydrogenase gene family and its diverse roles in response to multiple abiotic stress

Yanli Du[1,2†], Zhaoning Zhang[1†], Yanhua Gu[1], Weijia Li[1], Weiyu Wang[1], Xiankai Yuan[1], Yuxian Zhang[2,3], Ming Yuan[4], Jidao Du[1,2,5*] and Qiang Zhao[3,5*]

[1]Agricultural College, Heilongjiang Bayi Agricultural University, Daqing, Heilongjiang, China, [2]National Cereals Technology Engineering Research Center, Daqing, Heilongjiang, China, [3]Heilongjiang Bayi Agricultural University, Key Laboratory of Ministry of Agriculture and Rural Affairs of Soybean Mechanized Production, Daqing, Heilongjiang, China, [4]Qiqihar Branch of Heilongjiang Academy of Agricultural Sciences, Qiqihar, Heilongjiang, China, [5]Research Center of Saline and Alkali Land Improvement Engineering Technology in Heilongjiang Province, Daqing, Heilongjiang, China

Cytokinin oxidase/dehydrogenase (CKX) irreversibly degrades cytokinin, regulates growth and development, and helps plants to respond to environmental stress. Although the *CKX* gene has been well characterized in various plants, its role in soybean remains elusive. Therefore, in this study, the evolutionary relationship, chromosomal location, gene structure, motifs, *cis*-regulatory elements, collinearity, and gene expression patterns of *GmCKXs* were analyzed using RNA-seq, quantitative real-time PCR (qRT-PCR), and bioinformatics. We identified 18 *GmCKX* genes from the soybean genome and grouped them into five clades, each comprising members with similar gene structures and motifs. *Cis*-acting elements involved in hormones, resistance, and physiological metabolism were detected in the promoter regions of *GmCKXs*. Synteny analysis indicated that segmental duplication events contributed to the expansion of the soybean CKX family. The expression profiling of the *GmCKXs* genes using qRT-PCR showed tissue-specific expression patterns. The RNA-seq analysis also indicated that *GmCKXs* play an important role in response to salt and drought stresses at the seedling stage. The responses of the genes to salt, drought, synthetic cytokinin 6-benzyl aminopurine (6-BA), and the auxin indole-3-acetic acid (IAA) at the germination stage were further evaluated by qRT-PCR. Specifically, the *GmCKX14* gene was downregulated in the roots and the radicles at the germination stage. The hormones 6-BA and IAA repressed the expression levels of *GmCKX1*, *GmCKX6*, and *GmCKX9* genes but upregulated the expression levels of *GmCKX10* and *GmCKX18* genes. The three abiotic stresses also decreased the zeatin content in soybean radicle but enhanced the activity of the CKX enzymes. Conversely, the 6-BA and IAA treatments enhanced the CKX enzymes' activity but reduced the zeatin content in the radicles. This study, therefore, provides a reference for the functional analysis of *GmCKXs* in soybean in response to abiotic stresses.

KEYWORDS

abiotic stress, CKX gene family, soybean, expression analysis, growth and development

1 Introduction

Soybean (*Glycine max* L.) is an important food and oil crop worldwide, with its seed oil accounting for approximately 30% of the global vegetable oil consumption (Zhan et al., 2020). Soybean seeds contain various substances beneficial to human health, which have been proven important in preventing and treating cancer, atherosclerosis, osteoporosis, and coronary heart disease (Malenčić et al., 2012; Zhang et al., 2014). Soybeans also have a variety of industrial uses and are considered a potential crop for biodiesel production (Woyann et al., 2019). In the US, 30% of printed matter uses soybean ink, and many city buses have switched to an environmentally friendly blend of soybean oil and diesel (Frederick et al., 2001). Brazil, the world's top soybean producer, reportedly produced 124.5 million tons during 2019–2020 (USDA, 2020). Furthermore, soybean is one of the most common crops in arid and semi-arid areas where its growth and yield are easily affected by various abiotic stresses (Silvente et al., 2012; Gavili et al., 2019). Therefore, there is a need to explore the molecular mechanism involved in the soybean response to abiotic stress.

Cytokinin (CTK) is an important plant hormone that regulates many plant growth and development processes. CTKs naturally occurring in plants are derivatives of adenine and contain an isoprene-derived side chain or an aromatic side chain at the N^6 end, called isoprenoid CTKs and aromatic CTKs (Figure S1), respectively (Sakakibara, 2006). In plants, CTKs are distributed mainly in the dividing cells of stem and root tips, immature and germinating seeds, and growing fruits, promoting cell division and regulating differentiation. In tissue culture, the high ratio of CTKs to auxin benefits shoot differentiation, while a low ratio promotes root differentiation. The phytohormone also delays protein and chlorophyll degradation and plays a role in response to biological and abiotic stresses (Hwang et al., 2012; Jameson & Song, 2016; Wybouw & De Rybel, 2019).

Cytokinin oxidation/dehydrogenase (CKX) enzymes are encoded by a family of CKX genes that can specifically degrade unsaturated side chains at the N^6 position in CTKs and catalyze their irreversible degradation (Brownlee et al., 1975; McGaw & Horgan, 1983; Schmülling et al., 2003). Multiple gene families encoding CKX proteins (Werner et al., 2006) have been identified, and their evolution has been extensively studied in *Arabidopsis thaliana* (Werner et al., 2003), *Oryza sativa* (Ashikari et al., 2005; Rong et al., 2022), *Nicotiana tabacum* (Rong et al., 2022), *Zea mays* (Zalabák et al., 2014), *Triticum aestivum* (Mameaux et al., 2012), *Brassica rapa* (Mameaux et al., 2012), *Brassica napus* (Liu et al., 2018), *Brassica oleracea* (Zhu et al., 2022), and *Vitis vinifera* (Yu et al., 2021). The *CKX* genes also have several other functions in plants. For example, compared with the wild type, *atckx3/ckx5* double mutant showed increased stem apex meristem and silique number of *Arabidopsis* (Bartrina et al., 2011), while the inhibition of the expression of *OsCKX2* promotes rice growth by increasing its tiller number and yield (Gao et al., 2014; Yeh et al., 2015). The *OsCKX11* gene also regulates leaf senescence and grain number and coordinates the source–sink relationship in rice (Zhang et al., 2021). The *CKX* genes are also involved in plant responses to various biological and abiotic stresses. For example, suppressing the expression of the *CKXs* gene can significantly enhance *Arabidopsis* resistance against *Verticillium* wilt and fungal infection (Reusche et al., 2013). *Bol020547*, *Bol028392*, and *Bol045724* are important in determining cabbage (*B. oleracea* var. *capitata*) tolerance to *Plasmodiophora brassicae* (Zhu et al., 2022). In maize, most *CKX* genes were upregulated under salt stress (Vyroubalová et al., 2009); the overexpression of *CKXs* genes also enhanced *Arabidopsis* and tobacco tolerance to drought, salt, and abscisic acid stress (Nishiyama et al., 2011; Werner et al., 2011).

In this study, the *CKX* gene family in the whole soybean genome was identified and analyzed by bioinformatics techniques. The gene structure, chromosome distribution, *cis*-regulatory elements, gene replication, collinearity, and spatiotemporal expression patterns of the *GmCKX* genes were further analyzed. In addition, the key *GmCKX* genes that respond to salt, drought, salt combined with drought stress, 6-benzylaminopurine (6-BA), and indole-3-acetic acid (IAA) were screened. The results of this study lay the foundation for the study of *GmCKXs* gene function and provide important information for elucidating the evolutionary roles of *CKXs*.

2 Materials and methods

2.1 Identification and analysis of the *GmCKX* genes

The information of the reference genome and annotated proteins of soybean (*Glycine max* Wm82.a2.v1) was obtained from Ensembl Plants (http://plants.ensembl.org/index.html). The hidden Markov model (HMM) profile (http://hmmer.janelia.org/) and the Pfam database (http://pfam-legacy.xfam.org/) were used to screen candidate GmCKX proteins (PF01565 and PF09265). The CKXs protein sequence files were obtained from Ensembl Plants database (http://plants.ensembl.org/index). The InterPro (http://www.ebi.ac.uk/interpro/) (Finn et al., 2017) and SMART (http://smart.embl-heidelberg.de/) (Letunic et al., 2015; Han et al., 2019) software were used to further confirm the reliability of the CKX domain prediction. Then, the integrity of CKX domains was confirmed by Prosite (http://prosite.expast.org/) and WoLF PSORT (http://wolfpsort.hgc.jp/). All identified *GmCKX* genes were mapped according to their reference genome and named according to their locations on the chromosome using TBtools (Chen et al., 2020a). The CKX protein sequences derived from *Arabidopsis*, maize (*Z. mays* L.), and rice (*O. sativa* L.) were obtained from Ensembl Plants by searching CKX domains and used for phylogenetic analysis. A phylogenetic tree was constructed using the maximum likelihood method with 1,000 bootstrap replicates and the JTT+G model by MEGA X (version X-10.1.8, Mega Limited, Auckland, New Zealand).

The exon–intron structure of *GmCKX* genes was analyzed by the GSDS platform (http://gsds.cbi.pku.edu.cn/) (Guo, 2007). Gene-wise (Birney et al., 2004) was used to detect the correspondence between DNA and protein sequences. Then, the CKX domain coordinates in the protein sequence were converted to the coordinates in the nucleotide sequence using in-house perl script. The conserved motifs of CKX proteins were analyzed using MEME tool (http://meme.nbcr.net/meme/) (Bailey et al., 2009) with the following parameters: the motif length set at 10–50 amino acids and E value < $1e^{-20}$. The upstream regions (1,500 bp) of *GmCKX* genes were extracted and used as the gene promoter sequence. The *cis*-regulatory elements were analyzed by the PlantCare database (https://bioinformatics.psb.ugent.be/webtools/plantcare/html/). The Multiple Collinearity Scan toolkit (MCScanX) was used to analyze the synteny and collinearity of *GmCKX* genes (Wang et al., 2012). Subsequently, the collinearity of the duplicated genes was visualized by Circos software (version 0.69) (Krzywinski et al., 2009). The expression data of *GmCKXs* in different tissues came from the Phytozome database (https://phytozome-next.jgi.doe.gov) and Soybean ePF Browser database (http://bar.utoronto.ca/efpsoybean/cgi-bin/efpWeb.cgi), respectively. The heatmap was generated using TBtools (Chen et al., 2020a).

2.2 Plant materials and treatments

The soybean seeds (Heike68) were obtained from the National Coarse Cereals Engineering Research Center, Daqing, Heilongjiang, China. The surface-sterilized soybean seeds were placed on a petri dish measuring 9 cm in diameter and incubated in the dark at 28°C until germination, which was indicated by the emergence of radicles. After 5 days of germination under distilled water treatment (CK), samples of soybean cotyledons, radicles, and hypocotyls were harvested, frozen in liquid nitrogen for 5 min, and then stored at −80°C for tissue-specific expression analysis of *GmCKXs* using quantitative real-time PCR (qRT-PCR). After 4 days of germination, we selected seedlings with consistent growth to explore the response of their *GmCKXs* to different abiotic stress. The experimental treatments consisted of seedlings exposed to 150 mM NaCl (SS, simulated salt stress), 20% (W/V) PEG 6000 (D, simulated drought stress), 150 mM NaCl +20% (W/V) PEG 6000 (SS+D), 10 μM IAA, and 10 μM 6-BA. The seedlings were exposed to treatments as described previously (Liu et al., 2018), with those treated with distilled water (CK) alone as controls. We then incubated the treated seedlings at 28°C for 24 h in the dark, harvested radicle samples from the treated seedlings in liquid nitrogen, and then stored them at −80°C before further use.

Average-sized soybean seedlings without disease symptoms or insect spots were selected and sown in a polypropylene pot (upper diameter = 21 cm, lower diameter = 15 cm, and height = 19 cm). The pots were filled with peat-soil mixed with vermiculite at a volume ratio of 3:1 and pH 7.0 and maintained in a controlled environmental chamber with a light regime of 16 h/8 h (light/dark) and relative humidity of 50%–55% at 28 ± 2°C until the V1 stage. The seedlings were thinned to three per pot to obtain uniform seedlings and then treated with 50 ml of each CK (control), 75 mM NaCl (SS), 20% (W/V) PEG 6000 (D), and SS+D. After 5 days of treatment, the soybean root and shoot tissues were separated and collected in liquid nitrogen for 5 min, then stored at −80°C for RNA extraction and transcriptome analysis. The soybean seedlings in the same pot/petri dish were considered one experimental unit. All experiments were repeated three times.

2.3 RNA extraction, transcriptome analysis, and gene expression by qRT-PCR

Total RNA of soybean root and leaf samples were extracted using the Trizol reagent (Invitrogen, CA, USA), and their quality and purity were checked using the NanoDrop 2000 (Thermo Fisher Scientific, Wilmington, DE). The RNA Nano 6000 Assay Kit of the Agilent Bioanalyzer 2100 system (Agilent Technologies, CA, USA) was used to detect RNA integrity. The sequencing libraries were constructed by Biomarker Technologies Corporation (Beijing, China) on the Illumina HiSeq2500 as recommended by the manufacturer. After deleting the low-quality bases, the clean reads were mapped to the soybean genome (*Glycine max* Wm82.a2.v1). The differentially expressed genes (DEGs) with an adjusted *p*-value < 0.01 found by DESeq2 and FDR < 0.01 were assigned as differentially expressed.

The single-stranded cDNA of soybean seedling samples was synthesized using a 5 × HiScript SuperMix II according to the manufacturer's (Vazyme, Nanjing, China) instructions. The *GmCKXs* primers (Table S1) were designed with Primer 5.0 (Primer, Canada). Soybean *TUBULIN A* (NM_001250372) and *ACTIN* (NM_001289231) were used as the internal control genes. The qRT-PCR reaction was conducted using SYBR qPCR Master Mix (Vazyme, Nanjing, China) and run using the Roche Cycler 480II system (Roche, Roche Diagnostics, Switzerland). Relative expression levels for each CKX gene were calculated using the operational formula $2^{-\Delta\Delta Ct}$ (Livak and Schmittgen, 2001). Three technical replicates and three biological replicates were performed for each reaction for each sample in this study.

2.4 Determination of CKX enzyme activity and zeatin content

The CKX enzyme activity of samples was detected using the ELISA kit (10894, Meibiao, Jiangsu, China) according to the instructions. The zeatin content was determined using high-performance liquid chromatography (HPLC-MS/MS) (AB SCIEX, ShimadzuLc-20AD, AB5500 Massachusetts, USA) at the Customs Quality Inspection Center (Dalian, Liaoning, China).

2.5 Statistical analysis

Results were analyzed using one-way analysis of variance (ANOVA) and the Duncan's multiple range tests within the SPSS

19.0 (SPSS Inc., Chicago, IL, United States). Differences in values were considered statistically significant at $p < 0.05$.

3 Results

3.1 Identification and physicochemical property analysis of GmCKX genes in soybean

A total of 18 *GmCKX* genes were identified according to the result of an HMM profile, InterPro, and SMART analysis (Table 1). The results showed that the 18 GmCKX proteins contained amino acids (aa) ranging from 320 in *GmCKX06* to 552 aa in *GmCKX09*, with the lowest isoelectric point (IP) in *GmCKX11* (4.95) and the highest IP of 9.12 in *GmCKX07* and a low molecular weight (MW) of 35,800.26 Da in *GmCKX06* and a high MW of 62,281.19 Da in *GmCKX09*.

As shown in Figure 1, 18 *GmCKX* genes were unevenly distributed on 11 chromosomes: one gene on chromosome 3 (5.56% of the total), two genes on chromosome 4 (11.11% of the total), one gene on chromosome 6 (5.56% of the total), four genes on chromosome 9 (22.22% of the total), one gene on chromosome 11 (5.56% of the total), one gene on chromosome 12 (5.56% of the total), two genes on chromosome 13 (11.11% of the total), one gene on chromosome 14 (5.56% of the total), one gene on chromosome 15 (5.56% of the total), three genes on chromosome 17 (16.66% of the total), and one gene on chromosome 19 (5.56% of the total). In addition, *GmCKXs* are mostly distributed at both ends of the chromosomes.

3.2 Phylogenetic analysis of GmCKX proteins

To understand the evolution and development of the *CKX* gene family members in different species, 7 AtCKX, 11 OsCKX, 13 ZmCKX, and the 18 GmCKX proteins were assessed in a phylogenetic tree (Figure 2). The different CKXs were divided into six major clades (I–VI), with those from soybean only distributed in five subfamilies. Among the GmCKX proteins, clade I contained eight proteins, namely, GmCKX5, GmCKX6, GmCKX7, GmCKX11, GmCKX12, GmCKX14, GmCKX15, and GmCKX16; clade II had GmCKX2 and GmCKX4; clade III contained GmCKX3, GmCKX13, and GmCKX17; clade IV had GmCKX1 and GmCKX18; and clade V contained GmCKX8, GmCKX9, and GmCKX10 (Figures 2, S2).

3.3 Conserved motifs and gene structure analysis

The online software MEME was used to analyze the conservative motifs of GmCKXs. A total of 10 conserved motifs were obtained from the 18 GmCKXs, designated as Motifs 1 to 10 (Figures 3A, B). The *GmCKX* members in the same subfamily had similar motif characteristics but differed among *GmCKXs* in other subfamilies. Most *GmCKX* members contained 10 motifs each, with *GmCKX6* found in clade I containing five motifs, *GmCKX2* and

TABLE 1 Molecular characteristics of *GmCKX* genes in soybean.

Gene name	Gene_id	Chr	Location	Protein length (aa)	Isoelectric point	Molecular weight (Da)
GmCKX01	*Glyma.03G133300*	3	34850820/34853963	545	6.71	61,052.75
GmCKX02	*Glyma.04G028900*	4	2346613/2353642	424	6.23	48,279.08
GmCKX03	*Glyma.04G055600*	4	4492866/4496940	422	5.12	47,785.35
GmCKX04	*Glyma.06G028900*	6	2262638/2269432	424	5.87	48,345.8
GmCKX05	*Glyma.09G063500*	9	6102365/6107334	527	6.74	59,275.84
GmCKX06	*Glyma.09G063700*	9	6120899/6127022	320	5.87	35,800.26
GmCKX07	*Glyma.09G063900*	9	6163747/6168342	546	9.12	61,883.72
GmCKX08	*Glyma.09G225400*	9	45006788/45009855	534	6.37	60,063.71
GmCKX09	*Glyma.11G149100*	11	11564330/11568822	552	6.79	62,281.19
GmCKX10	*Glyma.12G011400*	12	831481/834416	538	6.24	60,336.71
GmCKX11	*Glyma.13G104600*	13	21926847/21931738	524	4.95	58,804.63
GmCKX12	*Glyma.13G104700*	13	21935494/21939535	535	7.35	60,986.77
GmCKX13	*Glyma.14G099000*	14	9505502/9511512	513	5.81	57,182.32
GmCKX14	*Glyma.15G170300*	15	15411545/15416020	543	8.73	61,546.2
GmCKX15	*Glyma.17G054500*	17	4143438/4147686	535	6.42	60,945.76
GmCKX16	*Glyma.17G054600*	17	4151290/4156276	522	5.34	58,758.95
GmCKX17	*Glyma.17G225700*	17	37956646/37963562	496	5.85	55,658.43
GmCKX18	*Glyma.19G135100*	19	39630174/39633290	545	6.12	60,841.21

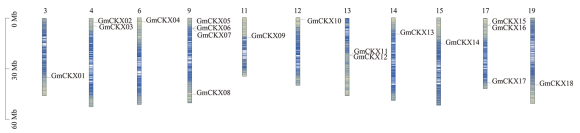


FIGURE 1
Localization of the *GmCKXs* on the soybean chromosomes. The chromosomal position of each *GmCKX* gene is shown on the corresponding chromosome from top to bottom according to the soybean genome. The blue line shows the gene density. The darker the color, the more dense the gene. The value on the *Y*-axis represents the position of the chromosome. The chromosome number is shown at the top of each bar.

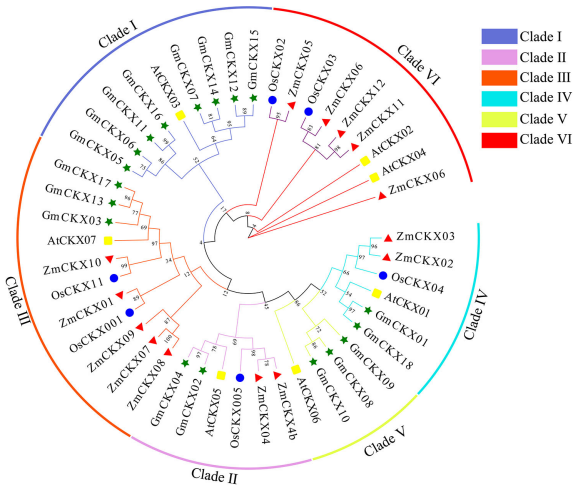


FIGURE 2
Phylogenetic analysis of CKX proteins in soybean, *Arabidopsis*, maize, and rice. The green, blue, red, and yellow circle represent soybean (*G. max* L.), *A. thaliana*, maize (*Z. mays* L.), and rice (*O. sativa* L.) respectively.

GmCKX4 (clade II), and GmCKX3 (clade III) containing eight motifs each. We analyzed the exon–intron structures of the *GmCKX* members on the GSDS website (Figure 3C) and found similar exon–intron structures for *GmCKXs* in the same subfamily, with differences among *GmCKXs* in different subfamilies. The number of exons in the 18 *GmCKX* genes ranged from four to seven, with the majority containing six exons each. The least number of exons was found in *GmCKX3*, while the highest number was found in *GmCKX8*.

3.4 Analysis of promoter *cis*-regulatory elements

To further study the regulatory mechanism of the *GmCKX* family in response to abiotic stress, the upstream 1.5-kb sequences of each of the 18 *GmCKXs* were extracted and used to analyze the *cis*-regulatory elements (Figure 4; Table S2). We identified 13 *cis*-regulatory elements and divided them into three groups: hormone-, resistance-, and physiological metabolism-related elements. The hormone-related elements consisted of the P-box, ABRE, TGA-element, TCA-element, GARE-motif, AuxRR-core, and TATC-box. The resistance-related elements included LTR, ARE, GC-motif, and MBS, while the physiological metabolism-related elements had only the MBSI and CAT-box.

3.5 Collinearity analysis

The origins of duplicates for *CKX* genes were detected by MCScanX and used to analyze the distribution and arrangement of its homologs within or between species. We identified six *GmCKX* duplicate gene pairs in the soybean genome, all characterized as segmental duplication events (Figure 5A; Table S3). Subsequently, we performed nonsynonymous and synonymous substitution ratio (Ka and Ks) analyses of duplicated genes to examine the driving forces of the soybean CKX gene family. The results showed that all six *GmCKX* gene pairs underwent purification selection with the Ka/Ks < 1. The collinearity analysis of *GmCKXs* with *Arabidopsis* was further used to explore the evolutionary mechanisms of the soybean CKX gene family. The results identified five orthologous gene pairs between soybean and *Arabidopsis* as collinear pairs, including *GmCKX1/AtCKX1*, *GmCKX2/AtCKX5*, *GmCKX5/AtCKX3*, *GmCKX6/AtCKX3*, and *GmCKX6/AtCKX3* (Figure 5B).

3.6 Expression profile analysis of *GmCKX* genes in soybean tissues

To explore the spatiotemporal expression patterns of soybean *GmCKX* genes, we compared the transcript abundances of all the 18 *GmCKX* genes using two publicly available RNA-Seq data from the Phytozome and Soybean ePF Browser database, respectively (Figures 6, 7). The Phytozome dataset contained root, root tip, lateral root, stem, shoot tip, leaf, flower, and nodules. In contrast, the Soybean ePF Browser dataset consisted of the root hair, shoot apical meristem (SAM), flower, green pods, leaf, nodule, root and root tip, and also the treatment and control root hair tissue after *Bradyrhizobium japonicum* infection at three different time points (Libault et al., 2010a; Libault et al., 2010b). Most *GmCKX* genes were preferentially expressed in more than one tissue, with *GmCKX7* and *GmCKX8* highly expressed and *GmCKX5* and *GmCKX6* genes lowly expressed or undetected in both two datasets. The data also showed that over 40% of the highly expressed *GmCKX* genes occurred in the floral organs of soybean.

3.7 Expression analysis of *GmCKX* genes in soybean seeds during the germination stage

Because seed germination is an important growth stage in the plant life cycle, we used qRT-PCR to investigate the expression of *GmCKXs* in the soybean seed's radicle, hypocotyl, and cotyledon during this stage (Figure 8). Some genes displayed tissue-specific expression. For example, *GmCKX3*, *GmCKX17*, and *GmCKX18*

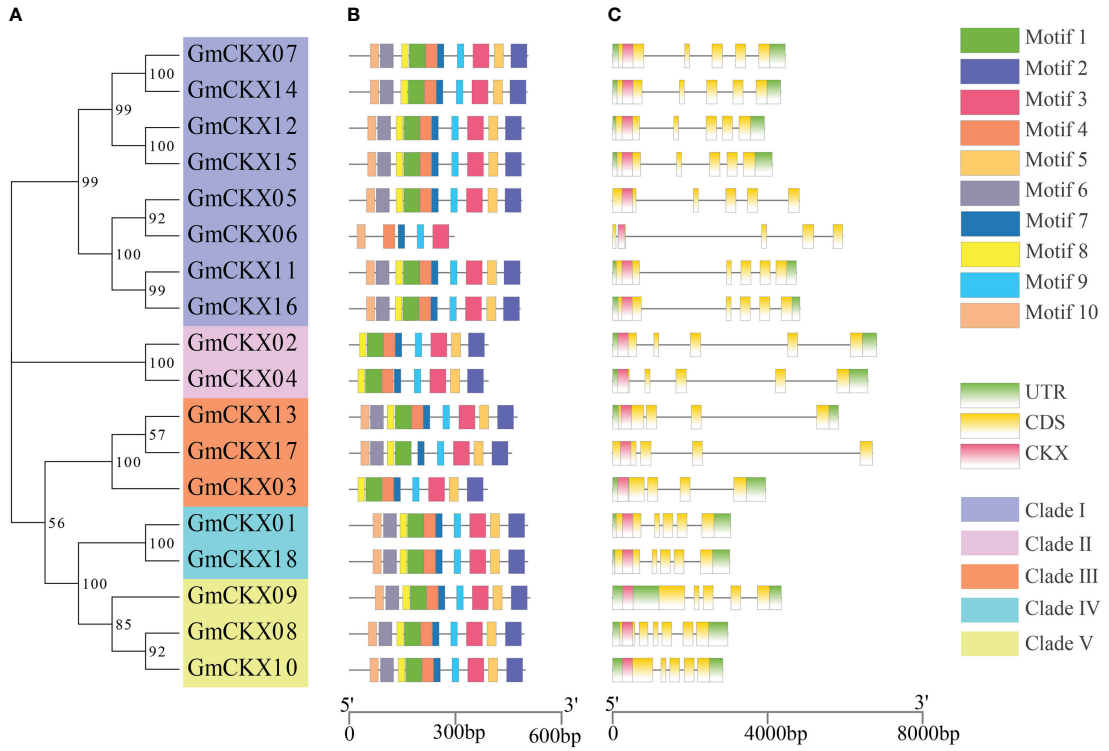


FIGURE 3
Gene structure and conserved motifs of the *GmCKXs*. **(A)** The phylogenetic classification of *GmCKXs*. **(B)** Conserved motif analysis of *GmCKXs*; different colored rectangles represented different motifs. **(C)** Gene structure analysis of *GmCKXs*. UTR regions (green rectangles), exons (yellow rectangles), CKX domains (pink rectangles), and introns (black lines).

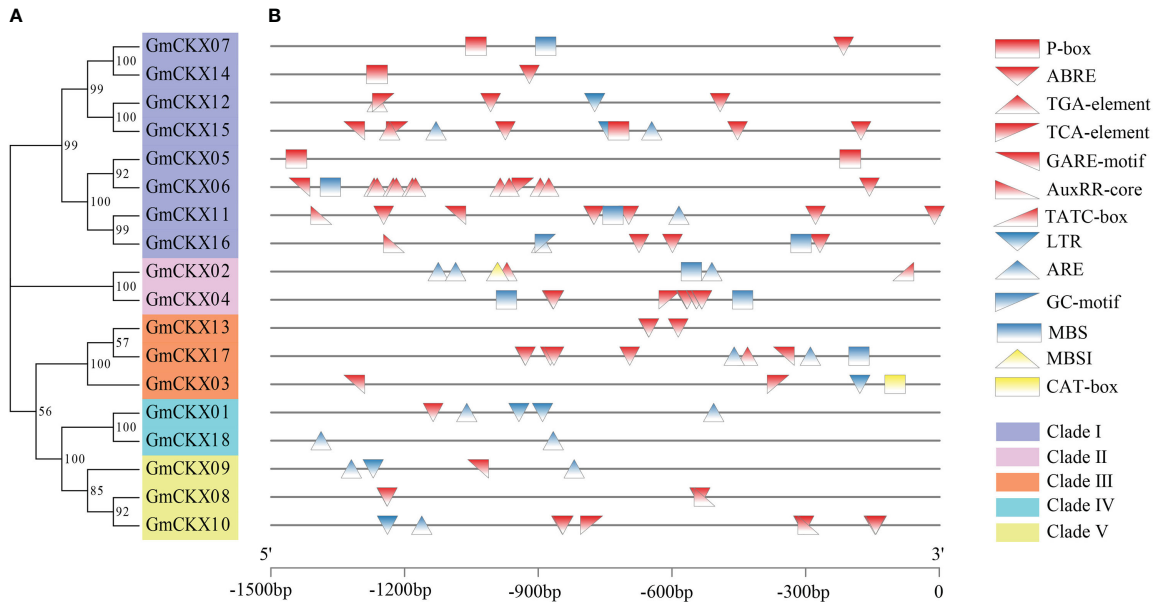


FIGURE 4
Promoter *cis*-regulatory element analysis of the *GmCKXs*. **(A)** The phylogenetic classification of *GmCKXs*. **(B)** *Cis*-element analysis of the promoter regions of *GmCKXs* genes. P-box: gibberellin-responsive element; ABRE: abscisic acid elements; TGA-element: auxin responsive element; TCA-element: salicylic acid elements; GARE-motif: gibberellin responsive element; AuxRR-core: auxin response promoter element; TATC-box: gibberellin responsive element; LTR: low-temperature responsiveness; ARE: anaerobic responsiveness; GC-motif: enhancer-like element involved in anoxic specific inducibility; MBS: drought stress inducibility element; MBSI: flavonoid biosynthesis regulation; CAT-box: meristem expression element.

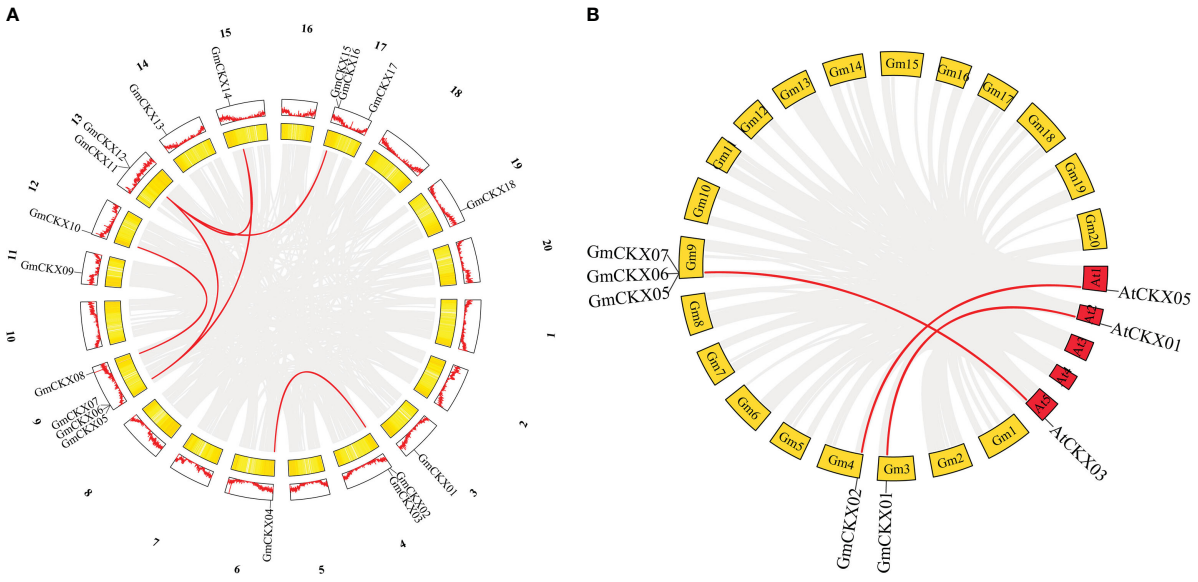


FIGURE 5
Collinearity analysis of the *GmCKXs*. **(A)** Duplicated gene pairs in soybean genome. Red lines indicate the duplication of *GmCKXs* gene pairs. **(B)** Collinearity analysis of *GmCKX* genes with *A. thaliana*. Red lines connect fragments of repeated gene pairs between soybean and *A. thaliana*.

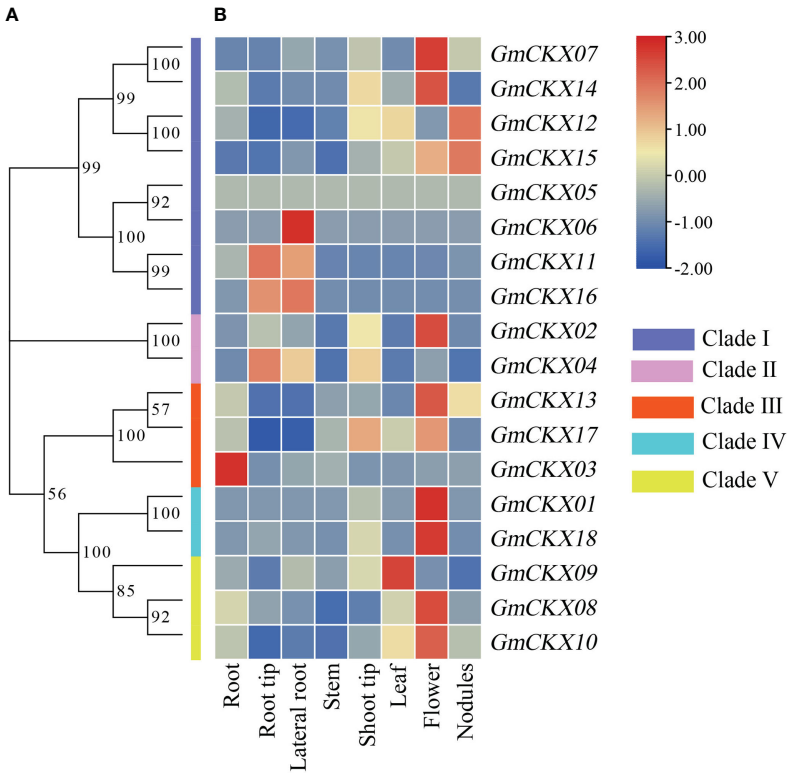


FIGURE 6
Heatmaps of the expression profiling of 18 *GmCKX* genes. **(A)** The phylogenetic classification of *GmCKXs*. **(B)** Expression profiling of *GmCKXs* in different tissues based on the Phytozome database. The color scale represents expression levels from high (red) to low (blue).

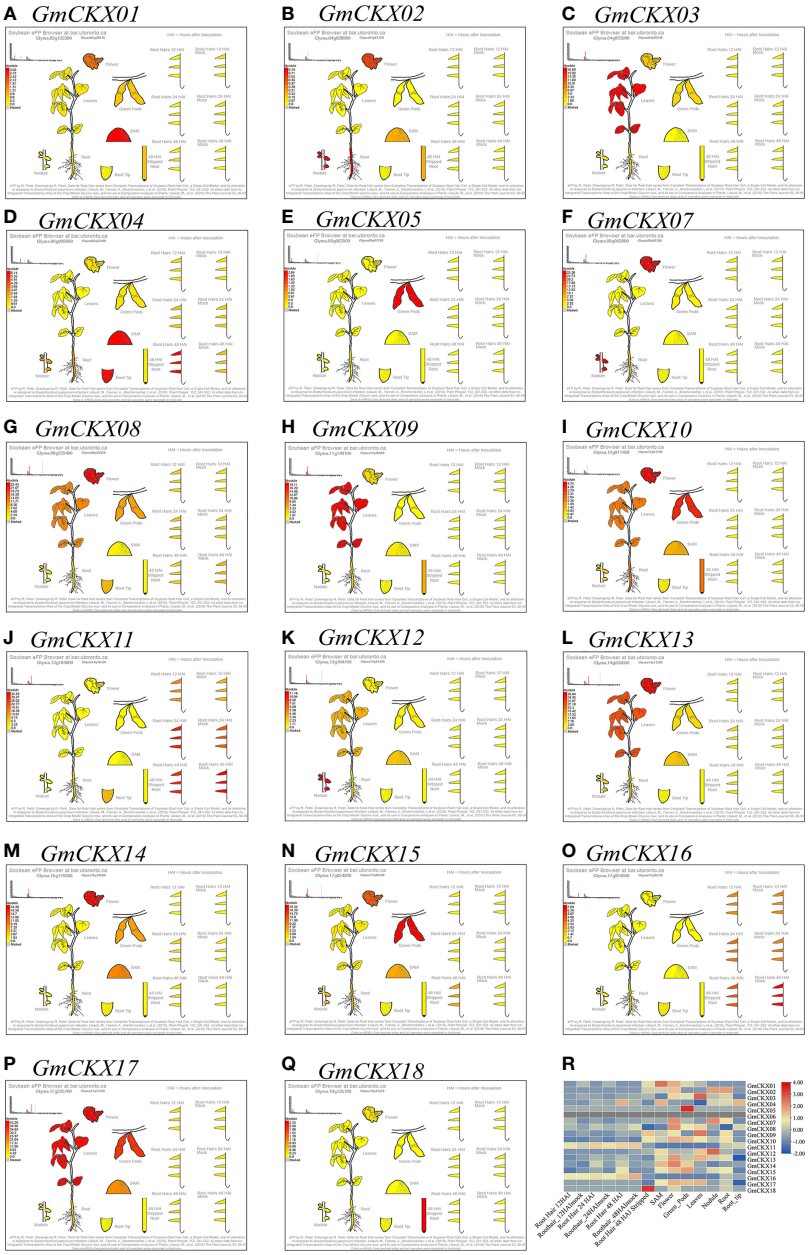


FIGURE 7
Expression profiling of *GmCKXs* in different tissues based on the Soybean eFP Browser database. **(A–Q)** Diagram showing the different soybean tissues. Red represents high expression and yellow represents low expression. **(R)** The heatmap were drawn by TBtools. Red represents high expression and blue represents low expression.

were found in the radicle; *GmCKX7, GmCKX9*, and *GmCKX10* were found in hypocotyl; and *GmCKX6*, *GmCKX8*, *GmCKX11*, and *GmCKX12* were found in the cotyledon.

3.8 Expression patterns of *GmCKXs* under abiotic stress

To explore the roles of specific *GmCKX* genes in response to different abiotic stresses, transcriptome expression patterns of all soybean *GmCKX* genes were analyzed in the leaves and roots of soybean seedlings under salt (SS), drought (D), and salt combined with drought stress (SS+D) (Figures 9A, B). The assembled gene dataset was deposited at the National Center for Biotechnology Information with the accession number PRJNA930177. Our data showed different expression profiles of *GmCKX* genes in different stress treatments and tissues. For example, compared to the control, three stress treatments significantly upregulated the expression level of *GmCKX13* in leaf and root but downregulated the expression level of *GmCKX3* and *GmCKX8*. The *GmCKX14* gene in soybean leaves was highly upregulated under the three stress treatments, while its expression level in the roots was significantly downregulated. The expression levels of *GmCKX9* in leaf and root were downregulated considerably under D and SS+D treatment, but SS treatment had no

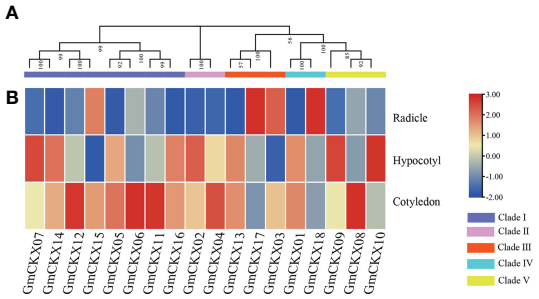


FIGURE 8
Expression pattern of soybean *GmCKX* genes in different tissues by qRT-PCR analysis during the germination stage. **(A)** The phylogenetic classification of *GmCKXs*. **(B)** Expression profiling of *GmCKXs* in radicle, hypocotyl, and cotyledon tissues. The color scale represents expression levels from high (red) to low (blue).

significant effects on its expression level. In the leaf, the *GmCKX15* gene was highly expressed under SS treatment but downregulated under D and SS+D treatment. At the same time, all three stress treatments significantly downregulated the expression of *GmCKX15* in the roots. The accuracy of transcriptome data was verified by qRT-PCR of six randomly selected *GmCKX* genes (Figures 9C–H).

To further investigate whether *GmCKXs* participate in response to the abiotic stresses during the germination stage, soybean seed radicles treated with SS, D, and SS+D were collected for qRT-PCR (Figure 10). We found that the expression of most *GmCKX* genes differed under different stress treatments. For example, compared with the control, the expression of *GmCKX1* and *GmCKX3* was significantly upregulated under SS treatment but significantly downregulated under D and SS+D treatments. The *GmCKX2* and *GmCKX8* were upregulated considerably under SS and D treatments but were significantly downregulated under SS+D treatment. All three stress treatments also significantly downregulated genes such as *GmCKX4*, *GmCKX7*, *GmCKX12*, *GmCKX13*, *GmCKX14*, *GmCKX15*, and *GmCKX17*, but significantly upregulated *GmCKX16*.

3.9 Hormone-induced patterns of expression of the *GmCKX* genes

We analyzed the relative expression level of *GmCKXs* in radicles treated with 6-BA and IAA using qRT-PCR to explore the hormone-induced patterns of expression of the *GmCKX* genes (Figure 11) and found differential expression of the *GmCKXs* under different hormone treatments. Compared to the control, the 6-BA and IAA significantly upregulated *GmCKX10* and *GmCKX18* genes but highly repressed the *GmCKX1*, *GmCKX6*, and *GmCKX9* genes. The *GmCKX2*, *GmCKX3*, *GmCKX7*, *GmCKX12*, *GmCKX13*, *GmCKX14*, *GmCKX15*, *GmCKX16*, and *GmCKX17* genes were upregulated considerably under 6-BA treatment but downregulated under IAA treatment. The *GmCKX4*, *GmCKX8*, and *GmCKX11* genes were significantly upregulated under 6-BA treatment but remained unaffected under IAA treatment. The IAA hormone also affected the expression of *the GmCKX5* gene by significantly upregulating it.

3.10 Abiotic stress and hormone-induced changes of CKX enzyme activity and zeatin content

We determined the zeatin content and CKX enzyme activity in soybean radicles under SS, D, SS+D, 6-BA, and IAA treatments to analyze the relationship between CKX enzyme activity and CTK

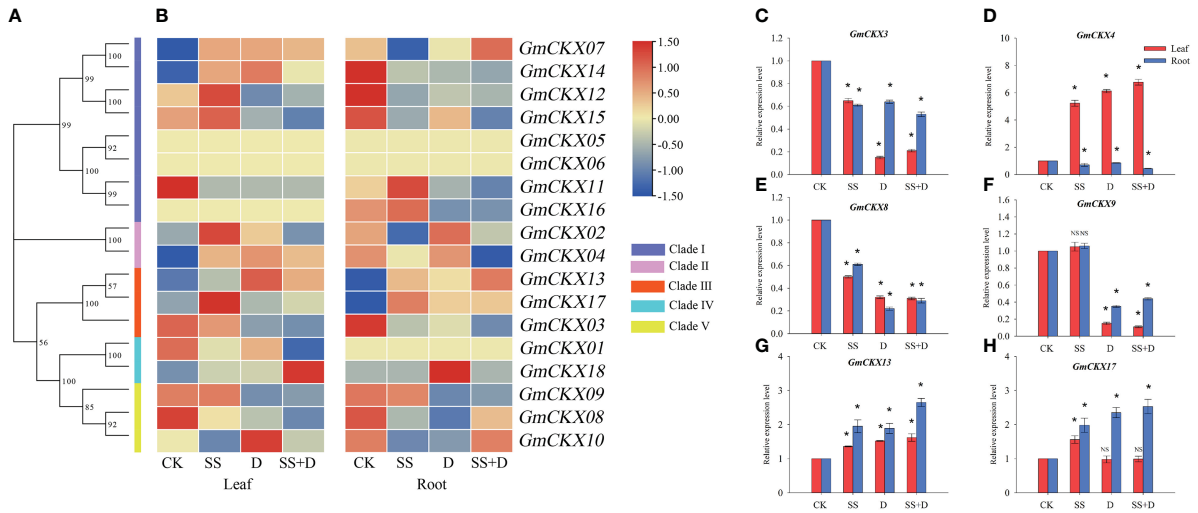


FIGURE 9
Expression profiles of soybean *GmCKXs* under different abiotic stresses at the seedling stage. **(A)** The phylogenetic classification of *GmCKXs*. **(B)** Expression profiling of *GmCKXs* in leaf and root under different abiotic stresses. CK, control condition; SS, salt stress; D, drought stress; SS+D, salt combined with drought stress. The color scale represents expression levels from high (red) to low (blue). **(C–H)** qRT-PCR analysis of six selected *GmCKX* genes in leaf and root under different abiotic stresses. Values are the means ± standard deviations of three replicates. * indicates significant difference between CK and treatment condition ($p < 0.05$). NS indicates no significant differences between CK and treatment condition.

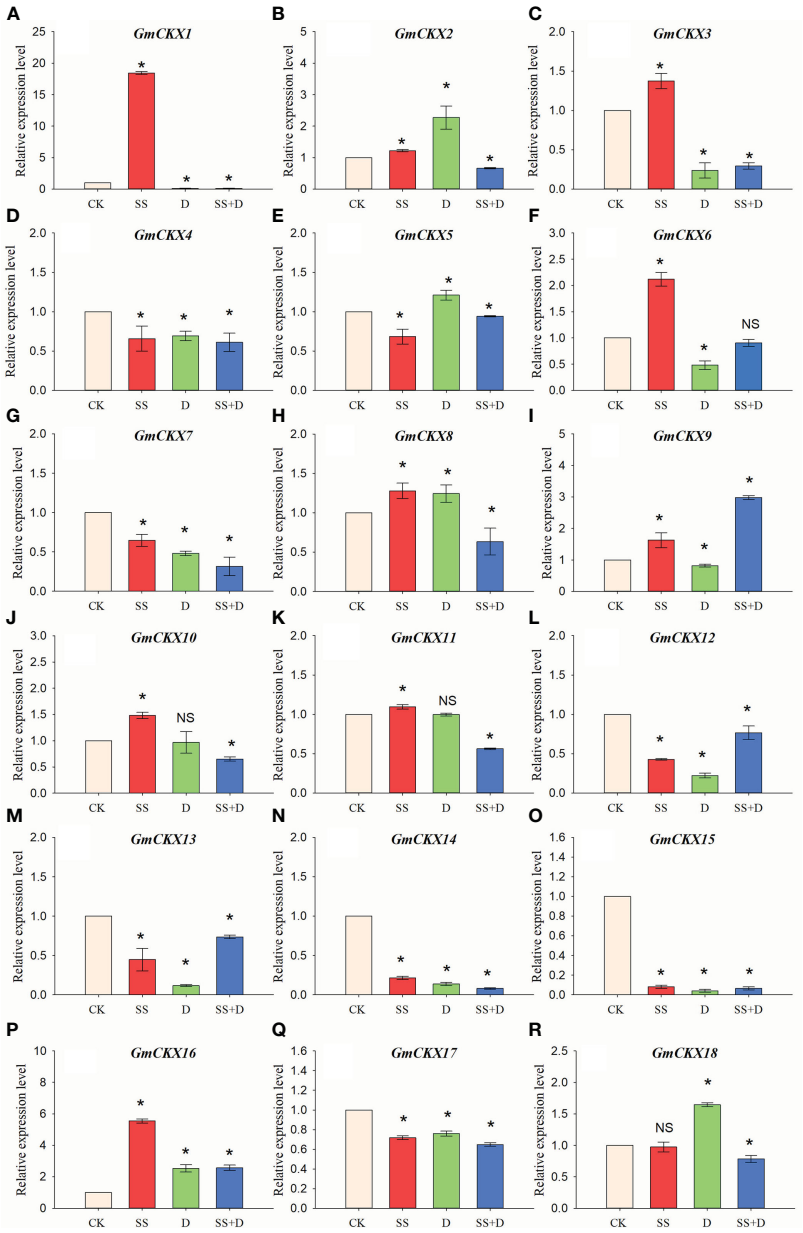


FIGURE 10
Relative expression levels of soybean *GmCKX* genes under SS, D, and SS+D treatments at the germination stage **(A–R)**. Values are the means ± standard deviations of three replicates. * indicates significant difference between CK and treatment condition ($p < 0.05$). NS indicates no significant differences between CK and treatment condition.

content under abiotic stress (Figure 12). Compared with the control, SS, D, and SS+D stress treatments significantly decreased zeatin content in soybean radicles by 32.02%, 44.2%, and 54.31%, respectively. In comparison, 6-BA and IAA treatments significantly increased the zeatin content in soybean radicles by 274.79% and 199.81%, respectively. Compared to the control, SS, D, and SS+D significantly increased CKX enzyme activity in soybean radicles by 32.9%, 39.35%, and 73.46%, respectively, while 6-BA and IAA significantly decreased CKX enzyme activity in soybean radicles by 38.33% and 26.9%, respectively.

4 Discussion

CTKs play an important role in numerous plant physiology processes, such as fatty acid biosynthesis in seed (Thien Nguyen et al., 2016), the development of plant floral organs and pod setting (Nonokawa et al., 2012), leaf senescence (Merewitz et al., 2010), and seed yield (Jameson and Song, 2016; Chen et al., 2020b). The hormone also helps the plants to respond to a variety of abiotic stresses, including drought (Hai et al., 2020), heat (Prerostova et al., 2020), and salt (Yu et al., 2022). Several members of the *CKX* gene

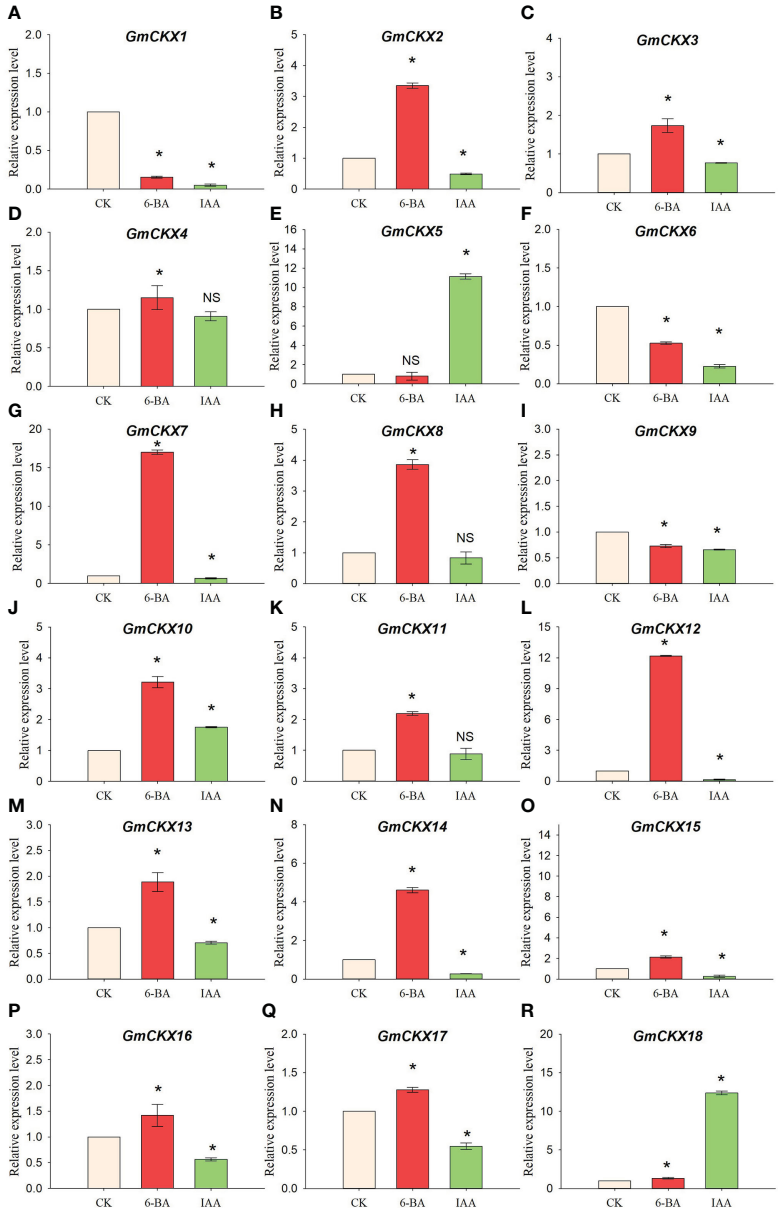


FIGURE 11
Relative expression levels of soybean *GmCKX* genes under IAA and 6-BA treatments **(A–R)**. Values are the means ± standard deviations of three replicates. * indicates significant difference between CK and treatment condition ($p < 0.05$). NS indicates no significant differences between CK and treatment condition.

family, which regulate the endogenous CKs, have been identified in various crops, including 11 members in rice (Mameaux et al., 2012), 13 in maize (Zalabák et al., 2014), 36 in cabbage (Zhu et al., 2022), 23 in oilseed rape (Liu et al., 2018), 5 in potato (Suttle et al., 2014), and 12 members in apple (Tan et al., 2018). In this study, we identified 18 *CKX* family members with complete domains in the entire genome of soybean, named *GmCKX1-GmCKX18*, according to their position on the chromosome. The variation in the number of *CKX* genes in different species is possibly due to genome evolution and replication, which causes the production of homologous genes and an increase in their numbers (Kaltenegger et al., 2018; Wang et al., 2021). We further divided the soybean CKXs into five subgroups (I–V) based on a phylogenetic tree containing *Arabidopsis* and soybean CKX protein sequences. Detailed molecular characterization of the *GmCKX* genes showed that CKX protein members had different physicochemical properties, such as protein length, pI, and MW, indicating a high diversity of these gene family members. These insights will help investigate the function of *GmCKXs*.

The exon–intron structure provides important information for exploring the evolutionary relationship of genes (Ahmad et al., 2019). In general, the number of exons plays an important role in gene evolution (Xu et al., 2012), and the number of introns determines the rate of gene expression (Jeffares et al., 2008; Shaul, 2017). Genes with similar exon–intron structures have similar gene functions. Therefore, gene function can be predicted by analyzing

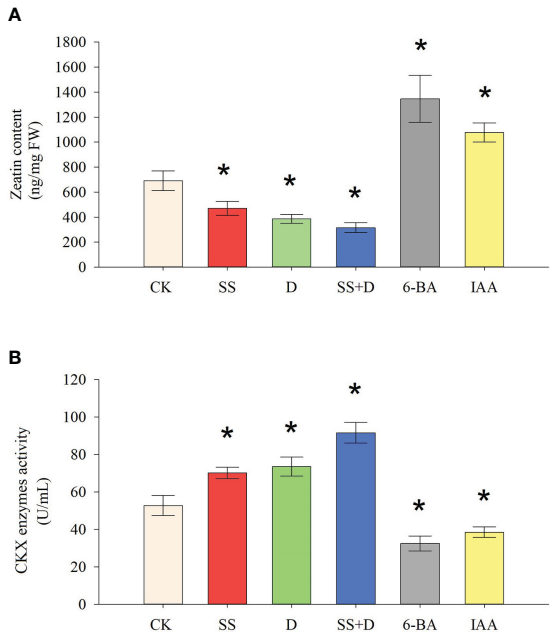


FIGURE 12
Zeatin content and CKX enzyme activity under different abiotic stresses and hormone treatments. **(A)** Zeatin content under SS, D, SS+D, 6-BA, and IAA treatments. **(B)** CKX enzyme activity under SS, D, SS+D, 6-BA, and IAA treatments. Values are the means ± standard deviations of three replicates. * indicates significant difference between CK and treatment condition ($p < 0.05$). NS indicates no significant differences between CK and treatment condition.

its structure (Li et al., 2019b). In the current study, most *GmCKX* genes in the same subfamilies had similar numbers of exons and introns, suggesting that they might have similar functions. However, we also found some *GmCKX* genes with different exon and intron numbers within the same subfamily, a phenomenon that has also been reported in oilseed rape (Liu et al., 2018) and cabbage (Zhu et al., 2022), which might be due to the functional diversity of genes throughout evolution.

The chromosomal localization analysis showed that the *GmCKX* genes were distributed non-homogeneously on chromosomes, showing a cluster distribution, which may be attributed to the non-uniform replication event of soybean chromosome fragments. Gene duplication is an important mechanism that promotes the expansion and diversification of gene families. Synteny analyses revealed that segmental duplication contributed to the expansion and diversification of the soybean *GmCKX* gene family. Similar results have been observed in maize (Gu et al., 2010) and Chinese cabbage (Liu et al., 2013). The study also showed that all six *GmCKX* gene pairs had Ka/Ks < 1, indicating that *GmCKXs* underwent purification selection under environmental stress, which is consistent with the results of previous studies (Yu et al., 2021; Zhu et al., 2022). Genome comparison is considered a relatively fast and effective method to study the potential characteristics and functions of genes (Lyons and Freeling, 2008). Therefore, the possible role of *CKX* homologous genes in the soybean can be inferred by analyzing the information on *CKX* genes in the model plant, such as *Arabidopsis*. This is supported by the location of five orthologous gene pairs in syntenic genomic regions between soybean and *Arabidopsis* genomes. For example, the *atCKX1* gene is expressed in root tissues and participates in lateral root formation (Chang et al., 2015). Ectopic expression of the *AtCKX1* gene in tobacco enhanced drought and heat stress tolerance (Macková et al., 2013). In *Arabidopsis*, *AtCKX3* and *AtCKX5* genes were expressed in reproductive meristems. The *ckx3 ckx5* double mutant could delay the differentiation of reproductive meristem cells and exhibit more and larger flower and silique numbers (Bartrina et al., 2011). Based on the reported function of the CKX gene (*AtCKX1*, *AtCKX3*, and *AtCKX5*) in *Arabidopsis*, we could predict the possible role of the soybean *GmCKX* genes. However, their functional roles need to be further confirmed in future reverse genetics studies.

Germination is the initial stage of soybean growth, which is also the most sensitive to environmental stress. The expression levels of genes in tissues and organs are closely related to their functions. We used two sets of public databases and qRT-PCR data from germination soybeans to detect the expression levels of the 18 *GmCKX* genes in different soybean tissues. The expression of the *GmCKX* genes differed among the soybean tissues, indicating that the *GmCKX* genes had different biological functions and were involved in soybean growth regulation and various tissue development processes. The expression patterns of individual *GmCKXs* in soybean were shown to be tissue and development specific. For example, *GmCKX7* and *GmCKX8* were highly expressed in all organs, while *GmCKX5* was mainly expressed in hypocotyl and cotyledon, and *GmCKX6* was highly expressed in cotyledon. These results are consistent with studies of *CKX* genes in other species (Song et al., 2012; Chen et al., 2020b; Yu et al., 2021).

CTK is a physiological hormone that widely exists in plants. As a key enzyme, which degrades endogenous CTK, CKX plays an important role in maintaining intracellular CTK homeostasis and for adaption to environmental stress (Vyroubalová et al., 2009; Le et al., 2012). *Cis*-acting elements play an important role in signal transduction and regulation of gene transcription initiation. Analysis of *cis*-acting elements in the promoters of *GmCKXs* demonstrated that the *GmCKXs* play a role in response to the hormone, plant growth, and biotic and abiotic stress responses. Similar findings were found in the *CKX* gene families in oilseed rape, maize, and *Arabidopsis*. For example, the overexpression of *MsCKX* genes in *Arabidopsis* exhibited stronger salt tolerance (Li et al., 2019a), while *ZmCKX1* was strongly induced by CTKs, abscisic acid, and abiotic stress in maize (Brugière et al., 2003). In oilseed rape, the expression level of *BnCKX7-1* was downregulated by the exogenous supply of 6-BA (Liu et al., 2018). In the current study, soybean *GmCKX* genes showed various roles in response to salt, drought, salt combined with drought stresses, 6-BA, and IAA. Our results demonstrate that each *GmCKX* gene is expressed differently in response to salt and drought stress and the exogenous supply of 6-BA and IAA hormones. The analysis of the *GmCKX* gene family at seedling and germination stages of soybean under salt, drought, and salt combined with drought stress showed that *GmCKX14* was downregulated both in root at the

seedling stage and in hypocotyl at the germination stage under the three abiotic stress treatments. The results of the evolutionary analysis showed that *GmCKX14* and *AtCKX3* were homologous genes. The overexpression of the *AtCKX3* gene increased the growth rate of primary roots and reduced the number of flowers in transgenic *Arabidopsis* (Werner et al., 2003). Transgenic tomatoes with overexpressed *AtCKX3* gene maintained the plants in a higher water state by reducing transpiration under drought treatment, thus enhancing their drought resistance (Farber et al., 2016). Our results suggested that soybean *GmCKX14* might be an important negative regulatory gene in abiotic stress such as salt and drought. The analysis of zeatin content and CKX enzyme activity in radicle under salt, drought, and salt combined with drought stress confirmed that abiotic stress enhanced CKX enzyme activity but reduced zeatin content.

In bread wheat, exogenous hormone treatment significantly induced *TuCKXs* gene expression within 3 h (Shoaib et al., 2019). Our results show that an exogenous supply of 6-BA and IAA could dramatically reduce the CKX enzyme activity of soybean radicle and enhance zeatin content. However, the soybean *GmCKXs* showed different expression patterns in response to 6-BA and IAA. After the exogenous supply of 6-BA and IAA, the *GmCKX1*, *GmCKX6*, and *GmCKX9* genes were all repressed, while the *GmCKX10* and *GmCKX18* genes were upregulated. Combined with physiological analysis results, the *GmCKX1*, *GmCKX6*, and *GmCKX9* genes could be used as positive regulatory factors, and *GmCKX10* and *GmCKX18* could be used as negative regulatory factors to participate in CK metabolism in response to exogenous 6-BA and IAA. However, this regulatory effect still requires further confirmation of its functional role in future reverse genetics studies. Our results also provide a reference for studying the function of the *CKX* gene under abiotic stress and hormonal regulation.

5 Conclusion

In this study, the 18 *GmCKX* genes were identified from the soybean genome, and their evolutionary relationship, chromosomal location, gene structure, motifs, *cis*-regulatory elements, collinearity, and gene expression patterns were analyzed by bioinformatics tools, RNA-seq, and qRT-PCR methods. The *GmCKXs* members were divided into five clades according to the phylogenetic tree. Synteny analyses revealed that the expansion of the *GmCKXs* gene family is mainly due to fragment replication. The analysis of Phytozome and Soybean ePF Browser databases and qRT-PCR showed that *GmCKX* genes had tissue-specific expression patterns. In addition, *GmCKXs* genes were differentially regulated in response to salt, drought, salt combined with drought stress, 6-BA, and IAA treatments. Expression of *GmCKX14* was downregulated both in root at the seedling stage and in radicle at the germination stage under salt, drought, and salt combined with drought stress treatments. Under 6-BA and IAA treatments, the expressions of *GmCKX1*, *GmCKX6*, and *GmCKX9* decreased, while the expressions of *GmCKX10* and *GmCKX18* increased. Finally, physiological analysis results showed that *GmCKX* genes could respond to abiotic stress and regulate the activity of CKX enzymes and the zeatin 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 below: https://www.ncbi.nlm.nih.gov/, PRJNA930177.

Author contributions

YD, ZZ, YG, JD, and QZ participated in the experimental design. YD, ZZ, WL, WW, XY, YZ, and MY performed material sampling, gene expression experiments, and physiology experiments. YD, ZZ, YG, and JD contributed to the data collection and data analysis. YD wrote the manuscript. ZZ, JD, and QZ revised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the National Key Research and Development Program (2021YFD1201603-3), the Heilongjiang Bayi Agricultural University Support Program for San Zong (ZRCQC202102 and ZRCQC202201), the Talent Introduction Project of Heilongjiang Bayi Agricultural University (XYB202006), the Postdoctoral General Project of Heilongjiang Province (LBH-Z21196), and the Innovative Research Project for Graduate Students of Heilongjiang Bayi Agricultural Reclamation University (YJSCX2022-Y01).

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1163219/full#supplementary-material

References

Ahmad, B., Zhang, S., Yao, J., Rahman, M. U., Hanif, M., Zhu, Y., et al. (2019). Genomic organization of the B3-domain transcription factor family in grapevine (*Vitis vinifera* l.) and expression during deed development in deedless and deeded cultivars. *Int. J. Mol. Sci.* 20, 4553. doi: 10.3390/ijms20184553

Ashikari, M., Sakakibara, H., Lin, S., Yamamoto, T., Takashi, T., Nishimura, A., et al. (2005). Cytokinin oxidase regulates rice grain production. *Science* 309, 741–745. doi: 10.1126/science.1113373

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. doi: 10.1093/nar/gkp335

Bartrina, I., Otto, E., Strnad, M., Werner, T., and Schmülling, T. (2011). Cytokinin regulates the activity of reproductive meristems, flower organ size, ovule formation, and thus seed yield in *Arabidopsis thaliana*. *Plant Cell* 23, 69–80. doi: 10.1105/tpc.110.079079

Birney, E., Clamp, M., and Durbin, R. (2004). Gene wise and genomewise. *Genome Res.* 14, 988–995. doi: 10.1101/gr.1865504

Brownlee, B. G., Hall, R. H., and Whitty, C. D. (1975). 3-Methyl-2-butenal: an enzymatic degradation product of the cytokinin, N^6-(Δ^2-isopentenyl) adenine. *Can. J. Biochem.* 53, 37–41. doi: 10.1139/o75-006

Brugière, N., Jiao, S., Hantke, S., Zinselmeier, C., Roessler, J. A., Niu, X., et al. (2003). Cytokinin oxidase gene expression in maize is localized to the vasculature, and is induced by cytokinins, abscisic acid, and abiotic stress. *Plant Physiol.* 132, 1228–1240. doi: 10.1104/pp.102.017707

Chang, L., Ramireddy, E., and Schmülling, T. (2015). Cytokinin as a positional cue regulating lateral root spacing in *Arabidopsis*. *J. Exp. Bot.* 66, 4759–4768. doi: 10.1093/jxb/erv252

Chen, C., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y., et al. (2020a). TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant* 13, 1194–1202. doi: 10.1016/j.molp.2020.06.009

Chen, L., Zhao, J., Song, J., and Jameson, P. E. (2020b). Cytokinin dehydrogenase: a genetic target for yield improvement in wheat. *Plant Biotechnol. J.* 18, 614–630. doi: 10.1111/pbi.13305

Farber, M., Attia, Z., and Weiss, D. (2016). Cytokinin activity increases stomatal density and transpiration rate in tomato. *J. Exp. Bot.* 67, 6351–6362. doi: 10.1093/jxb/erw398

Finn, R. D., Attwood, T. K., Babbitt, P. C., Bateman, A., Bork, P., Bridge, A. J., et al. (2017). InterPro in 2017–beyond protein family and domain annotations. *Nucleic Acids Res.* 45, D190–D199. doi: 10.1093/nar/gkw1107

Frederick, J. R., Camp, C. R., and Bauer, P. J. (2001). Drought-stress effects on branch and mainstem seed yield and yield components of determinate soybean. *Crop Sci.* 41, 759–763. doi: 10.2135/cropsci2001.413759x

Gao, S., Fang, J., Xu, F., Wang, W., Sun, X., Chu, J., et al. (2014). *CYTOKININ OXIDASE/DEHYDROGENASE4* integrates cytokinin and auxin signaling to control rice crown root formation. *Plant Physiol.* 165, 1035–1046. doi: 10.1104/pp.114.238584

Gavili, E., Moosavi, A. A., and Kamgar Haghighi, A. A. (2019). Does biochar mitigate the adverse effects of drought on the agronomic traits and yield components of soybean? *Ind. Crops Products* 128, 445–454. doi: 10.1016/j.indcrop.2018.11.047

Gu, R., Fu, J., Guo, S., Duan, F., Wang, Z., Mi, G., et al. (2010). Comparative expression and phylogenetic analysis of maize cytokinin dehydrogenase/oxidase (CKX) gene family. *J. Plant Growth Regul.* 29, 428–440. doi: 10.1007/s00344-010-9155-y

Guo, A.-Y. (2007). GSDS: a gene structure display server. *HEREDITAS* 29, 1023. doi: 10.1360/yc-007-1023

Hai, N. N., Chuong, N. N., Tu, N. H. C., Kisiala, A., Hoang, X. L. T., and Thao, N. P. (2020). Role and regulation of cytokinins in plant response to drought stress. *Plants* 9, 422. doi: 10.3390/plants9040422

Han, Z., Liu, Y., Deng, X., Liu, D., Liu, Y., Hu, Y., et al. (2019). Genome-wide identification and expression analysis of expansin gene family in common wheat (*Triticum aestivum* l.). *BMC Genomics* 20, 101. doi: 10.1186/s12864-019-5455-1

Hwang, I., Sheen, J., and Müller, B. (2012). Cytokinin signaling networks. *Annu. Rev. Plant Biol.* 63, 353–380. doi: 10.1146/annurev-arplant-042811-105503

Jameson, P. E., and Song, J. (2016). Cytokinin: a key driver of seed yield. *J. Exp. Bot.* 67, 593–606. doi: 10.1093/jxb/erv461

Jeffares, D. C., Penkett, C. J., and Bähler, J. (2008). Rapidly regulated genes are intron poor. *Trends Genet.* 24, 375–378. doi: 10.1016/j.tig.2008.05.006

Kaltenegger, E., Leng, S., and Heyl, A. (2018). The effects of repeated whole genome duplication events on the evolution of cytokinin signaling pathway. *BMC Evolutionary Biol.* 18, 76. doi: 10.1186/s12862-018-1153-x

Krzywinski, M., Schein, J., Birol, İ., Connors, J., Gascoyne, R., Horsman, D., et al. (2009). Circos: An information aesthetic for comparative genomics. *Genome Res.* 19, 1639–1645. doi: 10.1101/gr.092759.109

Le, D. T., Nishiyama, R., Watanabe, Y., Vankova, R., Tanaka, M., Seki, M., et al. (2012). Identification and expression analysis of cytokinin metabolic genes in soybean under normal and drought conditions in relation to cytokinin levels. *PLoS One* 7, e42411. doi: 10.1371/journal.pone.0042411

Letunic, I., Doerks, T., and Bork, P. (2015). SMART: recent updates, new developments and status in 2015. *Nucleic Acids Res.* 43, D257–D260. doi: 10.1093/nar/gku949

Li, S., An, Y., Hailati, S., Zhang, J., Cao, Y., Liu, Y., et al. (2019a). Overexpression of the cytokinin oxidase/dehydrogenase (CKX) from *Medicago sativa* enhanced salt stress tolerance of arabidopsis. *J. Plant Biol.* 62, 374–386. doi: 10.1007/s12374-019-0141-z

Li, Y., Chen, D., Luo, S., Zhu, Y., Jia, X., Duan, Y., et al. (2019b). Intron-mediated regulation of β-tubulin genes expression affects the sensitivity to carbendazim in *Fusarium graminearum*. *Curr. Genet.* 65, 1057–1069. doi: 10.1007/s00294-019-00960-4

Libault, M., Farmer, A., Brechenmacher, L., May, G. D., and Stacey, G. (2010a). Soybean root hairs: A valuable system to investigate plant biology at the cellular level. *Plant Signaling Behav.* 5, 419–421. doi: 10.4161/psb.5.4.11283

Libault, M., Farmer, A., Joshi, T., Takahashi, K., Langley, R. J., Franklin, L. D., et al. (2010b). An integrated transcriptome atlas of the crop model *Glycine max*, and its use in comparative analyses in plants. *Plant J.* 63, 86–99. doi: 10.1111/j.1365-313X.2010.04222.x

Liu, Z., Lv, Y., Zhang, M., Liu, Y., Kong, L., Zou, M., et al. (2013). Identification, expression, and comparative genomic analysis of the *IPT* and *CKX* gene families in Chinese cabbage (*Brassica rapa* ssp. *pekinensis*). *BMC Genomics* 14, 594. doi: 10.1186/1471-2164-14-594

Liu, P., Zhang, C., Ma, J.-Q., Zhang, L.-Y., Yang, B., Tang, X.-Y., et al. (2018). Genome-wide identification and expression profiling of cytokinin oxidase/dehydrogenase (CKX) genes reveal likely roles in pod development and stress responses in oilseed rape (*Brassica napus* l.). *Genes* 9, 168. doi: 10.3390/genes9030168

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262

Lyons, E., and Freeling, M. (2008). How to usefully compare homologous plant genes and chromosomes as DNA sequences: How to usefully compare plant genomes. *Plant J.* 53, 661–673. doi: 10.1111/j.1365-313X.2007.03326.x

Macková, H., Hronková, M., Dobrá, J., Turečková, V., Novák, O., Lubovská, Z., et al. (2013). Enhanced drought and heat stress tolerance of tobacco plants with ectopically enhanced cytokinin oxidase/dehydrogenase gene expression. *J. Exp. Bot.* 64, 2805–2815. doi: 10.1093/jxb/ert131

Malenčić, D., Cvejić, J., and Miladinović, J. (2012). Polyphenol content and antioxidant properties of colored soybean seeds from central Europe. *J. medicinal Food* 15 (1), 89–95. doi: 10.1093/jxb/ert131

Mameaux, S., Cockram, J., Thiel, T., Steuernagel, B., Stein, N., Taudien, S., et al. (2012). Molecular, phylogenetic and comparative genomic analysis of the *cytokinin oxidase/dehydrogenase* gene family in the poaceae. *Plant Biotechnol. J.* 10, 67–82. doi: 10.1111/j.1467-7652.2011.00645.x

McGaw, B. A., and Horgan, R. (1983). Cytokinin catabolism and cytokinin oxidase. *Phytochemistry* 22, 1103–1105. doi: 10.1016/0031-9422(83)80200-3

Merewitz, E. B., Gianfagna, T., and Huang, B. (2010). Effects of SAG12-ipt and HSP18.2-ipt expression on cytokinin production, root growth, and leaf senescence in creeping bentgrass exposed to drought stress. *J. Am. Soc. Hortic. Sci.* 135, 230–239. doi: 10.21273/JASHS.135.3.230

Nishiyama, R., Watanabe, Y., Fujita, Y., Le, D. T., Kojima, M., Werner, T., et al. (2011). Analysis of cytokinin mutants and regulation of cytokinin metabolic genes reveals important regulatory roles of cytokinins in drought, salt and abscisic acid responses, and abscisic acid biosynthesis. *Plant Cell* 23, 2169–2183. doi: 10.1105/tpc.111.087395

Nonokawa, K., Nakajima, T., Nakamura, T., and Kokubun, M. (2012). Effect of synthetic cytokinin application on pod setting of individual florets within raceme in soybean. *Plant Production Sci.* 15, 79–81. doi: 10.1626/pps.15.79

Prerostova, S., Dobrev, P. I., Kramna, B., Gaudinova, A., Knirsch, V., Spichal, L., et al. (2020). Heat acclimation and inhibition of cytokinin degradation positively affect heat stress tolerance of arabidopsis. *Front. Plant Sci.* 11, 87. doi: 10.3389/fpls.2020.00087

Reusche, M., Klásková, J., Thole, K., Truskina, J., Novák, O., Janz, D., et al. (2013). Stabilization of cytokinin levels enhances *Arabidopsis* resistance against *Verticillium longisporum*. *Mol. Plant-Microbe Interactions®* 26, 850–860. doi: 10.1094/MPMI-12-12-0287-R

Rong, C., Liu, Y., Chang, Z., Liu, Z., Ding, Y., and Ding, C. (2022). Cytokinin oxidase/dehydrogenase family genes exhibit functional divergence and overlap in rice growth and development, especially in control of tillering. *J. Exp. Bot.* 73, 3552–3568. doi: 10.1093/jxb/erac088

Sakakibara, H. (2006). CYTOKININS: activity, biosynthesis, and translocation. *Annu. Rev. Plant Biol.* 57, 431–449. doi: 10.1146/annurev.arplant.57.032905.105231

Schmülling, T., Werner, T., Riefler, M., Krupková, E., and Bartrina y Manns, I. (2003). Structure and function of cytokinin oxidase/dehydrogenase genes of maize, rice, arabidopsis and other species. *J. Plant Res.* 116, 241–252. doi: 10.1007/s10265-003-0096-4

Shaul, O. (2017). How introns enhance gene expression. *Int. J. Biochem. Cell Biol.* 91, 145–155. doi: 10.1016/j.biocel.2017.06.016

Shoaib, M., Yang, W., Shan, Q., Sajjad, M., and Zhang, A. (2019). Genome-wide identification and expression analysis of new cytokinin metabolic genes in bread wheat (*Triticum aestivum* l.). *PeerJ* 7, e6300. doi: 10.7717/peerj.6300

Silvente, S., Sobolev, A. P., and Lara, M. (2012). Metabolite adjustments in drought tolerant and sensitive soybean genotypes in response to water stress. *PLoS One* 7, e38554. doi: 10.1371/journal.pone.0038554

Song, J., Jiang, L., and Jameson, P. E. (2012). Co-Ordinate regulation of cytokinin gene family members during flag leaf and reproductive development in wheat. *BMC Plant Biol.* 12, 78. doi: 10.1186/1471-2229-12-78

Suttle, J. C., Huckle, L. L., Lu, S., and Knauber, D. C. (2014). Potato tuber cytokinin oxidase/dehydrogenase genes: Biochemical properties, activity, and expression during tuber dormancy progression. *J. Plant Physiol.* 171, 448–457. doi: 10.1016/j.jplph.2013.11.007

Tan, M., Li, G., Qi, S., Liu, X., Chen, X., Ma, J., et al. (2018). Identification and expression analysis of the IPT and CKX gene families during axillary bud outgrowth in apple (*Malus domestica* borkh.). *Gene* 651, 106–117. doi: 10.1016/j.gene.2018.01.101

Thien Nguyen, Q., Kisiala, A., Andreas, P., Neil Emery, R. J., and Narine, S. (2016). Soybean seed development: fatty acid and phytohormone metabolism and their interactions. *Curr. Genomics* 17, 241–260. doi: 10.2174/1389202917666160202220238

USDA (2020). United states department of agriculture (USDA) [WWW document]. *Foreing Agric. Service World Agric. Production*.

Vyroubalová, Š., Václavíková, K., Turečková, V., Novák, O., Šmehilová, M., Hluska, T., et al. (2009). Characterization of new maize genes putatively involved in cytokinin metabolism and their expression during osmotic stress in relation to cytokinin levels. *Plant Physiol.* 151, 433–447. doi: 10.1104/pp.109.142489

Wang, Y., Tang, H., DeBarry, J. D., Tan, X., Li, J., Wang, X., et al. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 40, e49–e49. doi: 10.1093/nar/gkr1293

Wang, C., Wang, H., Zhu, H., Ji, W., Hou, Y., Meng, Y., et al. (2021). Genome-wide identification and characterization of cytokinin oxidase/dehydrogenase family genes in *Medicago truncatula*. *J. Plant Physiol.* 256, 153308. doi: 10.1016/j.jplph.2020.153308

Werner, T., Köllmer, I., Bartrina, I., Holst, K., and Schmülling, T. (2006). New insights into the biology of cytokinin degradation. *Plant Biol.* 8, 371–381. doi: 10.1055/s-2006-923928

Werner, T., Motyka, V., Laucou, V., Smets, R., Van Onckelen, H., and Schmülling, T. (2003). Cytokinin-deficient transgenic arabidopsis plants show multiple developmental alterations indicating opposite functions of cytokinins in the regulation of shoot and root meristem activity. *Plant Cell* 15, 2532–2550. doi: 10.1105/tpc.014928

Werner, T., Nehnevajova, E., Köllmer, I., Novák, O., Strnad, M., Krämer, U., et al. (2011). Root-specific reduction of cytokinin causes enhanced root growth, drought tolerance, and leaf mineral enrichment in *Arabidopsis* and tobacco. *Plant Cell* 22, 3905–3920. doi: 10.1105/tpc.109.072694

Woyann, L. G., Meira, D., Zdziarski, A. D., Matei, G., Milioli, A. S., Rosa, A. C., et al. (2019). Multiple-trait selection of soybean for biodiesel production in Brazil. *Ind. Crops Products* 140, 111721. doi: 10.1016/j.indcrop.2019.111721

Wybouw, B., and De Rybel, B. (2019). Cytokinin - a developing story. *Trends Plant Sci.* 24, 177–185. doi: 10.1016/j.tplants.2018.10.012

Xu, G., Guo, C., Shan, H., and Kong, H. (2012). Divergence of duplicate genes in exon-intron structure. *Proc. Natl. Acad. Sci.* 109, 1187–1192. doi: 10.1073/pnas.1109047109

Yeh, S.-Y., Chen, H.-W., Ng, C.-Y., Lin, C.-Y., Tseng, T.-H., Li, W.-H., et al. (2015). Down-regulation of cytokinin oxidase 2 expression increases tiller number and improves rice yield. *Rice* 8, 36. doi: 10.1186/s12284-015-0070-5

Yu, Y., Li, Y., Yan, Z., and Duan, X. (2022). The role of cytokinins in plant under salt stress. *J. Plant Growth Regul.* 41, 2279–2291. doi: 10.1007/s00344-021-10441-z

Yu, K., Yu, Y., Bian, L., Ni, P., Ji, X., Guo, D., et al. (2021). Genome-wide identification of cytokinin oxidases/dehydrogenase (CKXs) in grape and expression during berry set. *Scientia Hortic.* 280, 109917. doi: 10.1016/j.scienta.2021.109917

Zalabák, D., Galuszka, P., Mrízová, K., Podlešáková, K., Gu, R., and Frébortová, J. (2014). Biochemical characterization of the maize cytokinin dehydrogenase family and cytokinin profiling in developing maize plantlets in relation to the expression of cytokinin dehydrogenase genes. *Plant Physiol. Biochem.* 74, 283–293. doi: 10.1016/j.plaphy.2013.11.020

Zhan, Y., Li, H., Sui, M., Zhao, X., Jing, Y., Luo, J., et al. (2020). Genome wide association mapping for tocopherol concentration in soybean seeds across multiple environments. *Ind. Crops Products* 154, 112674. doi: 10.1016/j.indcrop.2020.112674

Zhang, W., Peng, K., Cui, F., Wang, D., Zhao, J., Zhang, Y., et al. (2021). Cytokinin oxidase/dehydrogenase OsCKX11 coordinates source and sink relationship in rice by simultaneous regulation of leaf senescence and grain number. *Plant Biotechnol. J.* 19, 335–350. doi: 10.1111/pbi.13467

Zhang, S., Shi, Y., Zhang, S., Shang, W., Gao, X., and Wang, H. (2014). Whole soybean as probiotic lactic acid bacteria carrier food in solid-state fermentation. *Food Control* 41, 1–6. doi: 10.1016/j.foodcont.2013.12.026

Zhu, M., Wang, Y., Lu, S., Yang, L., Zhuang, M., Zhang, Y., et al. (2022). Genome-wide identification and analysis of cytokinin dehydrogenase/oxidase (CKX) family genes in *Brassica oleracea* l. reveals their involvement in response to *Plasmodiophora brassicae* infections. *Hortic. Plant J.* 8, 68–80. doi: 10.1016/j.hpj.2021.05.003

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 12 May 2023
DOI 10.3389/fpls.2023.1138988

OPEN ACCESS

EDITED BY
Jianghua Chen,
Key Laboratory of Tropical Plant Resource and Sustainable Use (CAS), China

REVIEWED BY
Uday Chand Jha,
Indian Institute of Pulses Research (ICAR), India
Juan M. Osorno,
North Dakota State University, United States
Xin Li,
Nanjing Agricultural University, China

*CORRESPONDENCE
Jing Yang
yangjing@zafu.edu.cn
Guojing Li
ligj@zaas.ac.cn

RECEIVED 06 January 2023
ACCEPTED 28 March 2023
PUBLISHED 12 May 2023

CITATION
Li M, Wu X, Wang B, Wu X, Wang Y, Wang J, Dong J, Wu J, Lu Z, Sun Y, Dong W, Yang J and Li G (2023) Genome-wide association analysis reveals the optimal genomic regions for pod size in bean. *Front. Plant Sci.* 14:1138988. doi: 10.3389/fpls.2023.1138988

Genome-wide association analysis reveals the optimal genomic regions for pod size in bean

Mao Li[1,2], Xinyi Wu[2,3,4], Baogen Wang[2,4], Xiaohua Wu[2,4], Ying Wang[2,4], Jian Wang[2,4], Junyang Dong[1,2], Jian Wu[1,2], Zhongfu Lu[2,4], Yuyan Sun[2,4], Wenqi Dong[2,4], Jing Yang[1]* and Guojing Li[2,4]*

[1]Key Laboratory of Quality and Safety Control for Subtropical Fruit and Vegetable, Ministry of Agriculture and Rural Affairs, Collaborative Innovation Center for Efficient and Green Production of Agriculture in Mountainous Areas of Zhejiang Province, College of Horticulture Science, Zhejiang A & F University, Hangzhou, China, [2]Institute of Vegetables, Zhejiang Academy of Agricultural Sciences, Hangzhou, China, [3]State Key Laboratory for Managing Biotic and Chemical Threats to the Quality and Safety of Agro-products, Zhejiang Academy of Agricultural Sciences, Hangzhou, China, [4]Key Laboratory of Vegetable Legumes Germplasm Enhancement and Molecular Breeding in Southern China (Co-construction by Ministry and Province), Ministry of Agriculture and Rural Affairs, Zhejiang Academy of Agricultural Sciences, Hangzhou, China

The snap bean is the most commonly grown vegetable legume worldwide, and its pod size is both an important yield and appearance quality trait. However, the improvement of pod size in snap beans grown in China has been largely hindered by a lack of information on the specific genes that determine pod size. In this study, we identified 88 snap bean accessions and evaluated their pod size traits. Through a genome-wide association study (GWAS), 57 single nucleotide polymorphisms (SNPs) significantly associated with pod size were detected. Candidate gene analysis showed that cytochrome P450 family genes, WRKY, and MYB transcription factors were the predominant candidate genes for pod development, and eight of these 26 candidate genes showed relatively higher expression patterns in flowers and young pods. A significant pod length (PL) SNP and a single pod weight (SPW) SNP were successfully converted into kompetitive allele-specific polymerase chain reaction (KASP) markers and validated in the panel. These results enhance our understanding of the genetic basis of pod size, and also provide genetic resources for the molecular breeding of pod size in snap beans.

KEYWORDS

pod size, GWAS, KASP marker, quality, snap bean

Introduction

The common bean (*Phaseolus vulgaris* L.) (2n = 2x = 22) is the most commonly grown grain and vegetable legume in the world. The common bean belongs to the *Leguminosae* family, which includes tropical legumes such as common beans (*P. vulgaris* L.), cowpeas [*Vigna unguiculata* (L.) Walp.], and soybeans (*Glycine max* L.), and cool season legumes such as peas (*Pisum sativum* L.) and faba beans (*Vicia faba* L.). It was recognized that the common bean was independently domesticated in two regions, now Mexico and South America, and formed the Middle American and Andean gene pools (Schmutz et al., 2014). The cultivated common bean comprised two main types, i.e., dry beans, used for food and fodder, and snap beans, used as vegetables. Dry beans are a valuable source of daily protein and calories globally and are essential to food and nutritional security in developing regions of the world, in particular in African and South American countries (Broughton et al., 2003). Snap bean is the most commonly grown vegetable legume in Europe, North America, and Asia because of its high nutritional content, which includes proteins, vitamins, and minerals (Delfini et al., 2021b).

For snap beans, pod size is not only an important yield-related trait, but also an important commercial trait that determines both consumer acceptance and whether or not good sale prices on the wholesale and retail markets are procured. Similar to other yield-related traits, pod size is a complex quantitative trait that is highly affected by the environment (Campa et al., 2018). Therefore, pod size improvement based on phenotype selection in traditional breeding is both a high-cost and a time-consuming process. To increase the precision and efficiency of selection for these complex traits, using genotype selection by way of marker-assisted tools is the most fundamental strategy, and exploiting the genes controlling the target traits, and developing the linked markers to the target genes, are key steps (Kamfwa et al., 2015). Previous studies have shown that most pod size traits such as pod length (PL), pod thickness (PT), and pod width (PW) displayed quantitative inheritance traits (Yuste-Lisbona et al., 2014; González et al., 2016; Hagerty et al., 2016; Murube et al., 2020), and some pod morphological traits were controlled by several major genes; for example, the cross-sectional shape trait was controlled by at least four genes (i.e., *Ea*, *Eb*, *Ia*, and *Ib*) (Leakey, 1988), the stringless trait was controlled by the *St* gene (Prakken, 1934), the twister trait was controlled by the *Tw* gene (Baggett and Kean, 1995), and the two genes *Da* and *Db* conferred the straight trait (Lamprecht, 1932; Lamprecht, 1947). During the past decade, many quantitative trait loci (QTLs) and SNP loci associated with pod size and pod morphological traits have been reported. Through linkage analysis, Yuste-Lisbona et al. (2014) identified six pod size QTLs on chromosomes Pv01, Pv04, and Pv11 in a recombinant inbred lines (RIL) population, comprising three for PL, two for PW, and one for PT. González et al. (2016) detected five QTLs for pod size on chromosomes Pv01, Pv02, and Pv09 using two RIL populations. Hagerty et al. (2016) mapped four pod size QTLs on chromosomes Pv04 and Pv09 using a RIL population derived from dry beans and snap beans. Murube et al. (2020) detected four QTLs for pod size on chromosomes Pv01, Pv02, Pv07, and Pv11 using two nested RIL populations. Interesting, all the studies have detected some QTLs with pleiotropic effects on PL, PW, and PT, although these were mapped on different chromosomes. Using genome-wide association mapping in a panel of 135 accessions, Gupta et al. (2020) detected two and nine simple sequence repeats (SSRs) associated with PL and PW, respectively. García-Fernández et al. (2021) detected 23 QTLs for PL, six for pod cross-section, and six for pod characters from 301 bean lines of the Spanish Diversity Panel.

The common bean was introduced to China in the 15th century, and the dry bean variety is now among the top six cereal substitute crops, with an annual production of 1.29 million tons. The snap bean is the most commonly grown legume vegetable, with an annual production of 17.99 million tons in China (FAO2020). Over 4,900 common bean accessions have been collected in China, showing notable diversity (Zhang et al., 2008). Wu et al. (2019) have detected one genomic region associated with pod height, 13 genomic regions associated with PL, and 59 genomic regions associated with PW using a diversity panel of 683 common bean accessions, in which most accessions belonged to dry bean. Because of the very limited pod size genes/QTLs reported in Chinese snap bean germplasms, the genetic improvement of pod size in snap bean breeding has been largely hindered in China. To solve this problem, in the present study, we aimed to conduct a genome-wide association study (GWAS) for pod size in a diversity panel of Chinese snap bean accessions, including landraces and cultivars. The results enhance our understanding of the genetic basis of pod size and facilitate the molecular breeding of pod size in snap beans.

Materials and methods

Plant materials and pod size evaluation

A diversity panel consisting of 88 snap bean accessions were used in this study. The 88 accessions included 62 landraces and 26 cultivars, of which 12 accessions belonged to the Andean gene pool (Wu et al., 2022). In a previous study, this panel was used to identify the significant genomic regions for bean rust (*Uromyces appendiculatus*) (Wu et al., 2022). All accessions were planted in the Haining Yangdu Scientific Research and Innovation Base of Zhejiang Academy of Agricultural Sciences in 2020 and 2021 using a completely randomized design with two replications. Each accession was sown in four plots on two columns of 1.5 m-wide beds, with 40 cm inter-row spacing and 75 cm individual spacing. Normal field management was used for plant growth and development over the 2 years.

For each accession, at least 10 representative immature pods were collected for pod size evaluation. These pods were grown for about 15 days after pollination, which was when they reached the commercial maturity stage and could be used only as vegetables. The PL was measured using a ruler, PW and PT were measured on the middle section of pods using a vernier caliper, and the single pod weight (SPW) was measured using a scale. The phenotype data of two replications on each trait were averaged. The data analyses, including frequency distributions, correlation analyses, and variance analyses, were conducted using Origin software (Version 2018). Multivariate analysis of variance (MANOVA) was conducted using SPSS software (v.22) under a general linear model (GLM).

Genotyping and population structure analysis

The genotype data of the 88 accessions were retrieved from the study by Wu et al. (2022). What follows is a brief summary of the procedure. All accessions were sequenced with 12 × genome coverage using Illumina re-sequencing, 652.97 Gb of data were generated, and the clean reads were aligned to the common bean reference genome *P. vulgaris* v2.1 (https://phytozome-next.jgi.doe.gov/info/Pvulgaris_v2_1) using Burrows-Wheeler Aligner (BWA) software (http://bio-bwa.sourceforge.net/). Subsequently, the aligned files were converted to binary alignment map (BAM) files using SAMtools software, and, finally, the valid BAM files were used for SNP calling using the genome analysis toolkit (GATK) "UnifiedGenotyper" function (http://www.broadinstitute.org/gatk/). A total of 20,175,784 SNPs were identified. The criteria of a missing call rate > 0.5 and minor allele frequency (MAF) < 0.05 were used to filter the original SNPs. Finally, 603,910 high-quality SNPs were retained for population structural analysis and GWAS.

GWAS

To detect the genomic regions associated with pod size, a GWAS was conducted on each trait using Tassel (v5.2.82) under a compressed mixed linear model (MLM) that accounted for population structure. The genetic effect of each SNP to the total phenotypic variation was calculated based on the marker R^2 values. The SNPs showing logarithm of the odds (LOD) values of ≥3.5 or 6.0 were regarded as significant SNPs. If two significant SNPs were located in the same linkage disequilibrium (LD) block (Delfini et al., 2021b; Valdisser et al., 2017; Wu et al., 2019), they were considered to represent the same QTL.

Synteny analysis of pod size QTLs

The reported pod size QTLs/genes in common bean were first identified through a literature search, and then the sequences of the linked markers to these genes were blasted against the common bean reference genome *P. vulgaris* v2.1 (https://phytozome-next.jgi.doe.gov/info/Pvulgaris_v2_1) to determine their physical position under an e-value cut-off of $1e^{-10}$. By comparing the physical positions of the known pod size QTLs/genes and new QTLs detected in this study, their syntenic relations were determined, and if only two QTLs/genes were located in a LD block they were considered to represent a single gene/locus.

Candidate genes mining and gene expression analysis

According to the estimated LD decay distance for common bean (Delfini et al., 2021b; Valdisser et al., 2017; Wu et al., 2019), the potential candidate genes residing in the 100 kb upstream and downstream of each SNP locus were investigated in the *P. vulgaris* v2.1 genome. Based on their annotation information, the genes with functions in developmental processes were considered as possible candidate genes. In addition, the gene expression data in different tissues were retrieved from *P. vulgaris* v2.1 and analyzed using Tbtools software (v1.098).

KASP markers development and validation

To facilitate the molecular breeding of pod size in the common bean, the sequence of the significant SNP was used to design kompetitive allele-specific polymerase chain reaction (KASP) markers using a local script. The KASP genotyping for the 88 accessions was performed on an IntelliQube workstation (LGC Genomics Ltd., Hoddesdon, UK) following the PCR process described in the study by Wu et al. (2021). The haplotype analysis was conducted using Origin software (Version 2018).

Results

Phenotypic variation of pod size traits

As shown in Table 1 and Figure 1, the pod size showed a high level of variation in the 88 common bean accessions over the 2 years. The PLs ranged from 10.22–41.82 cm and 9.64–24.15 cm in 2020 and 2021, respectively. The PWs ranged from 7.61–18.64 mm and 5.22–17.18 mm in 2020 and 2021, respectively. The PTs ranged from 5.76 mm to 10.62 mm in 2020, and 3.72 mm to 6.86 mm in 2021. The SPWs also demonstrated continuous variation, ranging from 4.60–20.60 g in 2020 and 4.40–16.70 g in 2021. PL and PT showed the largest and smallest variance values by environmental changes in individual years, respectively, indicating that PL displayed the highest variation on the population level. The analysis of variance (ANOVA) statistics indicated that all four traits were affected significantly, by environmental changes in individual years with significant differences presenting among all genotypes, excluding those for PT, and significant interaction between genotypes and environments displayed for PL and PW (Table S1), indicating that it may be inappropriate to combine the two years data for further analysis. All four pod size traits displayed a nearly normal distribution in the diversity panel (Figure 1). Among the four traits, PL showed a higher positive correlation with both SPW (0.872) and PT (0.671), and a higher positive correlation (0.749) was also found between PT and SPW (Figure 2).

Marker trait associations

In a previous study, the 88 accessions were genotyped by re-sequencing and used for detecting SNPs significantly associated with rust resistance (Wu et al., 2022). The results of principal component analysis (PCA), neighbor-joining tree analysis, and structural analysis all supported that these accessions could be

TABLE 1 The statistics of pod size traits in this study.

Trait according to year	Max	Min	Mean ± SE	Var	CV (%)
2021 PL (cm)	24.15	9.64	15.53 ± 0.35	10.99	21.35%
2020 PL (cm)	41.82	10.22	17.8 ± 0.54	24.39	27.75%
2021 PW (mm)	17.18	5.22	9.41 ± 0.27	6.43	26.97%
2020 PW (mm)	18.64	7.61	10.62 ± 0.26	5.72	22.53%
2021 PT (mm)	6.86	3.72	5.41 ± 0.06	0.36	11.21%
2020 PT (mm)	10.62	5.76	7.69 ± 0.09	0.76	11.34%
2021 SPW (g)	16.70	4.40	8.31 ± 0.23	4.72	26.17%
2020 SPW (g)	20.60	4.60	10.19 ± 0.29	7.03	26.04%

Max, maximum; Min, minimum; SE, standard error; Var, variance; CV, coefficient of variation; PL, pod length; PT, pod thickness; PW, pod width; SPW, single pod weight.

divided into two main clusters, which corresponded with the Middle American and Andean gene pools (Wu et al., 2022). To detect the pod size signals, a GWAS was conducted and a total of 1,241 SNPs with significant association with pod size were identified under the threshold value ≥ 3.5, which was distributed on all 11 chromosomes and accounted for 13.5%–59.1% of the phenotypic variation (Figure 3). It was also found that some SNPs formed clusters in different chromosomes and their adjacent distance ranged between 1 bp and hundreds of bps, indicating that some SNPs may represent a single pod size gene based on the 100 kb LD decay distance in common beans (Delfini et al., 2021b; Valdisser et al., 2017; Wu et al., 2019). After removing the redundant SNPs in an LD block and filtering the PW loci using a higher threshold value (≥ 6.0), a final 57 representative loci were reported (Table 2).

Of the significant SNPs found, nine were for PL, seven for PT, 22 for PW, and 19 for SPW (Table 2). The PL SNPs were distributed on five chromosomes (i.e., Pv01, Pv04, Pv05, Pv07, and Pv11) and accounted for 13.5%–23.9% of the phenotypic variation. The PW signals were from all 11 chromosomes, excluding chromosome 10, and accounted for 34.0%–55.7% of the phenotypic variation. The PT SNPs were located on chromosomes Pv01, Pv02, Pv03, Pv07, and Pv08, and accounted for 19.7%–30.8% of the phenotypic variation. The SPW SNPs were located on six chromosomes (i.e., Pv01, Pv02, Pv03, Pv06, Pv08, and Pv10), and accounted for 17.3%–38.3% of the phenotypic variation. A total of 16 significant SNPs were detected in both environments, and 15 were PW SNPs (Table 2), indicating that PW has greater heredity and was less affected by the environments.

Candidate genes analysis

A total of 618 predicated gene models were identified as the candidates for the 57 pod size SNPs, which belonged to multiple gene families, such as the cytochrome P450 protein family, major facilitator protein superfamily, and the MYB transcription factor family (Table S2). Among them, 26 gene models with annotated functions highly associated with plant growth and the development of 16 significant SNPs were identified (Table 3). Of which, 13 were cytochrome P450 proteins gene models which involve in the regulation of plant hormone metabolism and play important roles in plant growth and development (Xu et al., 2015). Among these 13 cytochrome P450 protein gene models, six (*Phvul.004G005700-Phvul.004G006100*) formed a cluster from 368,531 bp to 403,445 bp on chromosome Pv04, which has a 24.471 kb distance to the PL SNP Pv_0285901. We also identified two MYB transcription factor family gene models (*Phvul.008G094800, Phvul.010G075500*) and three WRKY protein gene models (*Phvul.008G090300, Phvul.002G265400, Phvul.002G266400*) that play crucial regulatory roles in plant developmental processes (Dubos et al., 2010; Wang et al., 2021). In addition, we identified other gene models, i.e., *Phvul.011G015300*, *Phvul.008G088100*, *Phvul.003G128800*, *Phvul.004G166000*, *Phvul.006G157800*, and *Phvul.006G159300*, involved in plant hormone biosynthesis and regulation that play a central role in controlling plant growth and development (Cui et al., 2010; Feng et al., 2013; Nibau et al., 2013; Zhao et al., 2015; Salem et al., 2017; Powers et al., 2019; Jing et al., 2022). The glycosyltransferase gene model *Phvul.005G091200* was proposed to be involved in secondary cell wall glucuronoxylan and/or pectin biosynthesis that increasing the stem growth in Populus (Biswal et al., 2015). The gene model *Phvul.003G072200* encodes an amino phospholipid ATPase, which complex forms an important part of the Golgi machinery required for secretory processes during plant development (Poulsen et al., 2008).

The results of an *in silico* analysis showed that 26 candidate gene models demonstrated a significantly different expression pattern in different tissues, and eight genes were found to be more abundantly expressed in flower buds, flowers, young pods, and green mature pods, indicating that they may be involved in the pathways of pod development (Figure 4). The genes with higher expression levels in young pods included the amino phospholipid ATPase family protein gene (*Phvul.003G072200*), the auxin response factor (*Phvul.003G128800*), cytochrome P450 gene (*Phvul.04G005400*, *Phvul.04G005800*), the hormone biosynthesis and regulation genes (*Phvul.011G015300* and *Phvul.006G159300*), the MYB transcription factor *Phvul.010G075500* and the WRKY gene *Phvul.008G090300*, suggesting that hormonal and nutritional signaling are also important mechanisms related to pod size control.

10.3389/fpls.2023.1138988

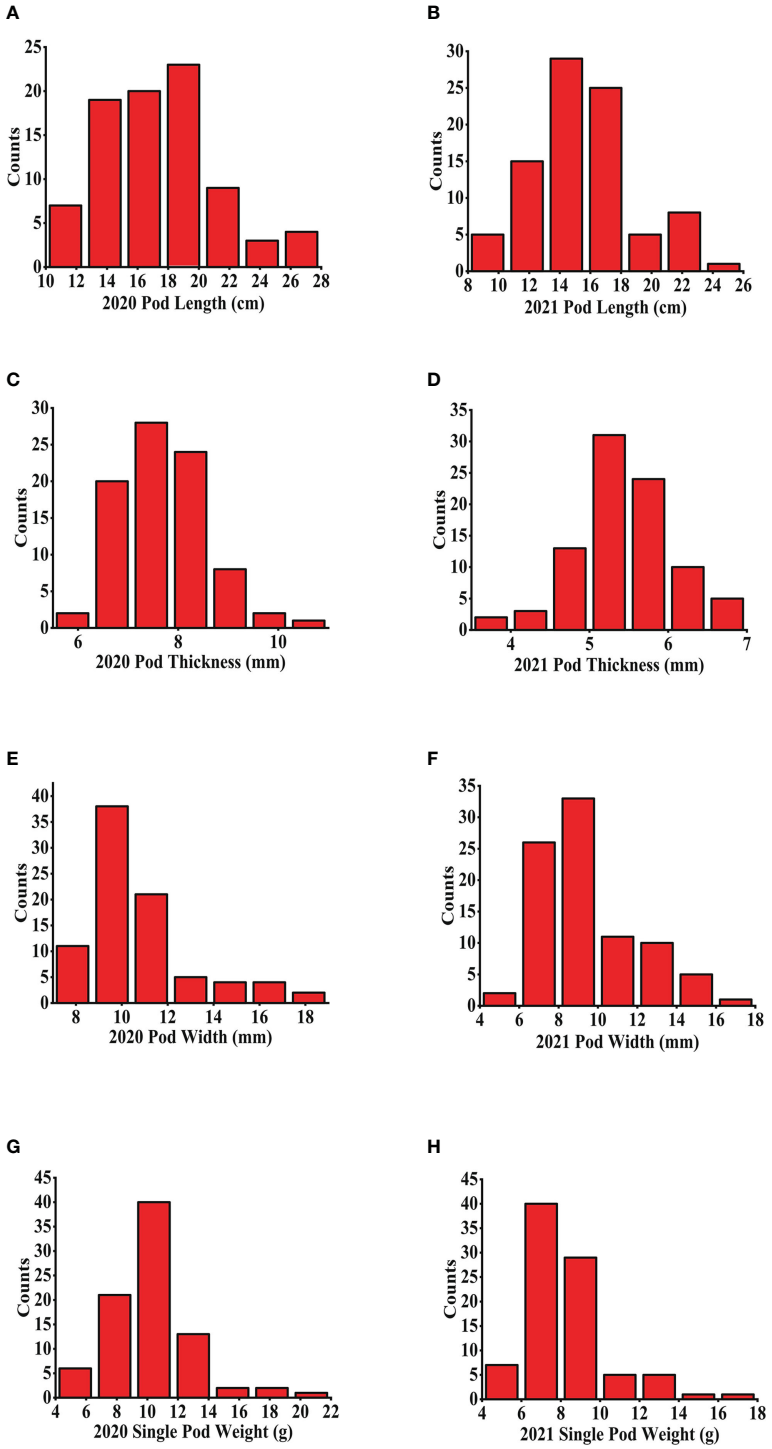


FIGURE 1
The frequency distribution of pod size over the 2 years. **(A, B)** PLs in 2020 and 2021, respectively; **(C, D)** PTs in 2020 and 2021, respectively; **(E, F)** PWs in 2020 and 2021, respectively; **(G, H)** SPWs in 2020 and 2021, respectively. PL, pod length; PT, pod thickness; SPW, single pod weight; PW, pow width.

Development of KASP markers for the important loci

To facilitate the molecular breeding of pod size, a PL SNP Pv_0026128 and a SPW SNP Pv_974939 were selected for conversion into PCR-based KASP markers (Table 4) and were successfully validated in the diversity panel (Figures 5A, E). The Pv_0026128 showed *CC* and *TT* genotypes for 88 accessions, and the haplotype Pv_0026128-CC showed a longer PL (Figures 5B–D; the average PL is 17.96 cm), which could increase the PL by 5.08 cm. The Pv_974939 displayed an AA and a GG haplotype with 60 and 25 accessions, respectively. The haplotype Pv_974939-GG had more

10.3389/fpls.2023.1138988

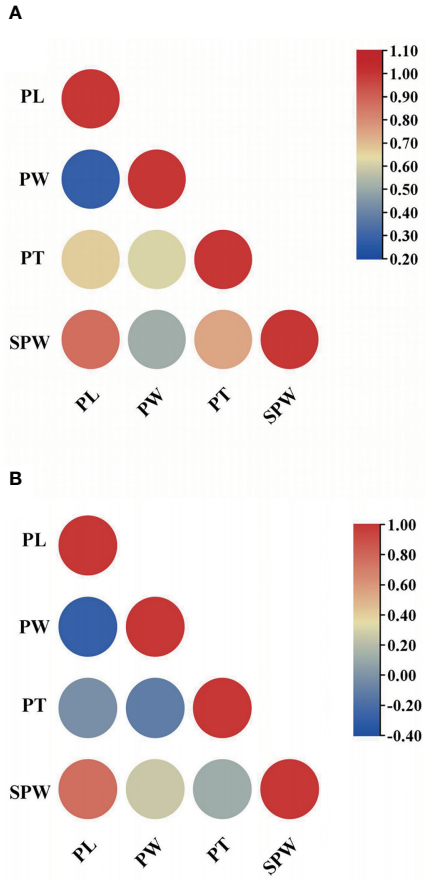


FIGURE 2
Correlations among pod size traits. **(A)** the correlation in 2020; **(B)** the correlation in 2021.

SPW than that of Pv_974939-AA (Figures 5F–H; the average SPW is 9.68 g), which increased the SPW by 1.84 g. These results indicate that the two markers could be used for PL and SPW selection in snap bean breeding.

Discussion

Snap bean is the most commonly grown vegetable legume for human consumption. It is believed that snap beans are dry beans that were selectively bred after the Columbian Exchange and that have since been subject to intense selective breeding separately in Europe and China, the purpose of which has been to enhance such traits as low pod wall fiber and thick pod walls (Myers and Baggett, 1999; Zhang et al., 2008; Wallace et al., 2018). Indeed, pod size has long been of interest to breeders because it is an important yield and appearance quality trait. The improvement of pod size in snap bean requires a fundamental understanding of the genetic basis of pod size, and the associations between other pod morphological traits and QTL/genes have been identified in studies of foreign bean germplasms (Yuste-Lisbona et al., 2014; González et al., 2016; Hagerty et al., 2016; Gupta et al., 2020; Murube et al., 2020; García-Fernández et al., 2021). However, even though China is the largest snap bean producer in the world, investigations into the genetic architecture of pod size in Chinese snap bean germplasms are lacking. In the present study, a diversity panel consisting of 88 Chinese snap bean germplasms that has previously been used to identify rust resistance signals (Wu et al., 2022) was evaluated for pod size and a total of 57 SNPs significantly associated with pod size were detected using GWAS. The panel size is relatively small; however, population structure analysis showed that it could be divided into the Andean and Middle American gene pools that are well established in common bean, a finding that is consistent with the population division of a larger Chinese common bean in which 683 accessions were identified (Wu et al., 2019). In addition, pod size traits such as PL and PW showed higher levels of heritability (0.89 and 0.91, respectively) (García-Fernández et al., 2021), indicating the stronger resolution of these traits for gene mapping. Therefore, based on the above two conditions, we deemed it practicable to use a small panel to investigate the genetic architecture related to pod size. Among this panel, 12 accessions belonged to the Andean gene pool; this ratio is lower than that of the American snap bean germplasms, in which most snap bean accessions belonged to the Andean gene pool (Song et al., 2015). This is not surprising given that China is the secondary center of diversity for common bean (Zhang et al., 2008), and snap bean has been widely cultivated in diverse agro-ecosystems, with environments ranging from subtropical to temperate and from sea level to 3,000 meters above sea level (masl). Therefore, snap beans were possibly selectively bred from both the Andean and Middle American gene pools by people in diverse areas, and natural or artificial inter-gene pool introgression may have also led to this shift. In fact, previous studies have suggested that snap beans may have originated from more than one gene pool (Wallace et al., 2018). In any case, our results indicate the optimal genes and phenotypes that could be used in breeding lines, especially in China.

The LD decay distance varied in different common bean collections using different Brazilian common bean diversity panels. Valdisser et al. (2017) found that the LD extension estimate was 395 kb for the Andean gene pool and 130 kb for the Middle American gene pool, and Delfini et al., (2021b) calculated that the LD decay value is 296 kb. Using a Chinese common bean diversity panel consisting of 683 accessions, Wu et al. (2019) found that the average LD decay distance between SNPs was 107 kb across the genome. In our previous study, the LD decay distance was found to be 800 kb (Wu et al., 2022). Because of the relatively small population size in this study, however, we regarded an LD decay distance of 800 kb as being unrepresentative. Considering that all the 88 accessions belonged to Chinese landraces and cultivars, which may bear a much closer resemblance to the diversity panel used in a previous study by Wu et al. (2019), we finally decided on a decay distance of ± 100 kb LD to identify the overlapped SNPs and search the candidate gene models for each significant SNP.

According to the ±100 kb LD, it was found that a PT SNP Pv_0201428 (chr02: 48,802,744) has a 92.501 kb distance to the PW SNP Pv_0203965 (chr02: 48,895,245), indicating that they may represent a single locus. This pleiotropic effect on PW and PT was also investigated in a previous study, in which a pod width QTL PWI4 and a pod thickness QTL PT4.1 were mapped on the same interval (Yuste-Lisbona et al., 2014). We also found that these two

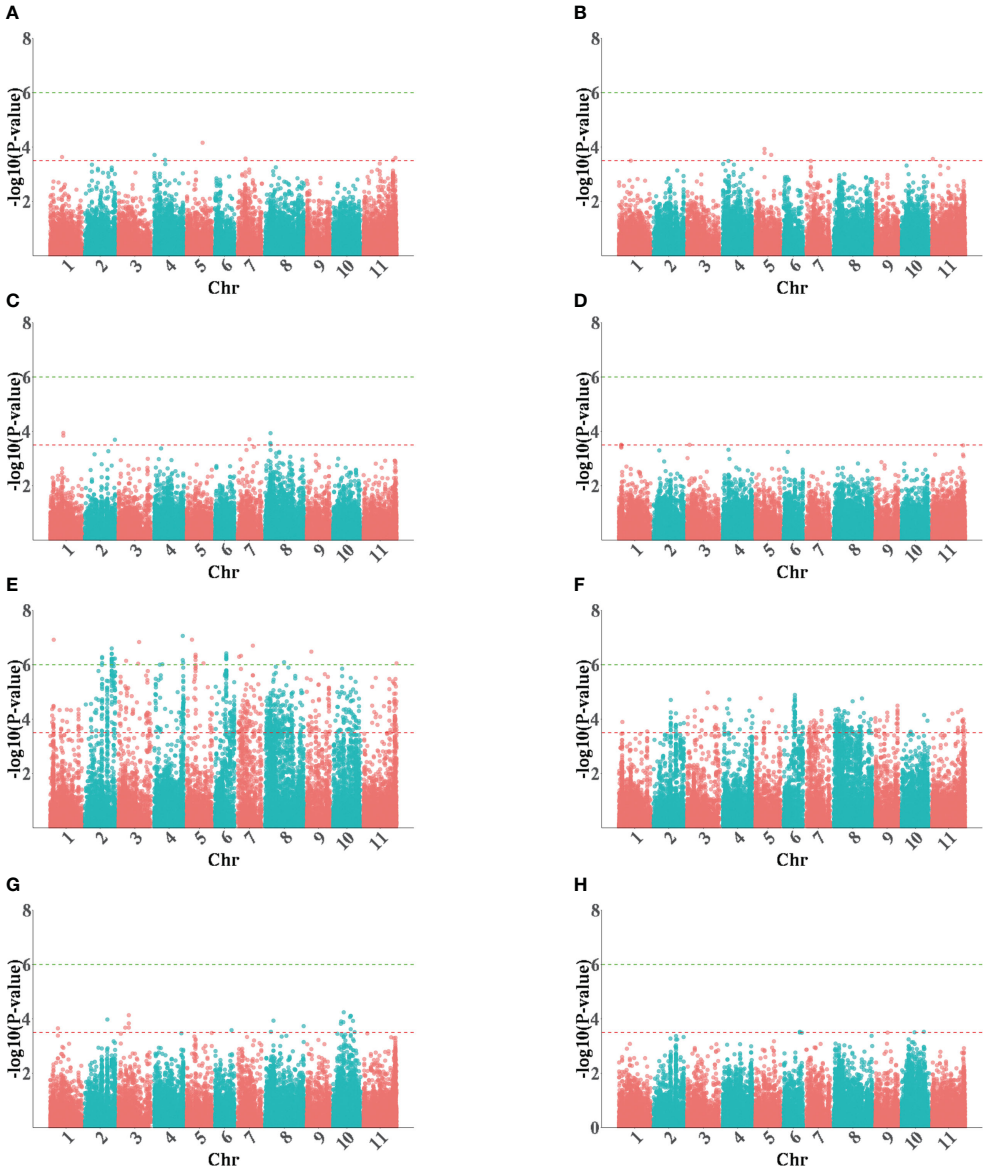


FIGURE 3
The Manhattan plots of GWAS for the pod size in 2020 (left) and 2021(right). **(A, B)** PLs; **(C, D)** PTs; **(E, F)** PWs; **(G, H)** SPWs. PL, pod length; PT, pod thickness; SPW, single pod weight; PW, pow width.

SNPs were located in a QTLs enrichment interval from 48,634,684 bp to 49,605,168 bp on chromosome 2, where five pod-related QTLs including a pod color QTL (PodLCol2_49.4), a PL QTL (PL2.2XC), two numbers of seeds per pod QTLs (NSP02_48.7 and NSP2XC), and a pod perimeter QTL (E-PP2XB) (Murube et al., 2020; García-Fernández et al., 2021), were detected. In addition, we found that a PW SNP Pv_0519622 and a SPW SNP Pv_0237496 were located in the intervals of the PW QTL PWI6 and the QTL NSP3 for the number of seeds per pod (Murube et al., 2020), respectively, suggesting that pod morphological trait QTLs often demonstrate pleiotropic effects, or that they are clustered, or closely linked, in the genome.

As a fruit organ, pod development is a complex process that is regulated by plant hormones and transcription factors (Tang et al., 2020). Cytochrome P450 family genes WRKY and MYB transcription factors were the predominant candidate genes for pod size SNPs detected in this study, which were also the main candidates for pod morphological QTLs in previous study (García-Fernández et al., 2021). For the PL SNP Pv_0285901, a total of eight cytochrome P450 genes were identified as its candidates, six of which formed a cluster and two genes, *Phvul.004G005400* and *Phvul.004G005800*, showed higher expression levels in flowers and young pods (Table 3; Figure 4). The cytochrome P450 genes *BnaC7.ROT3* and *BnaA9.CYP78A9* both mediate silique length by affecting cell elongation in rapeseed through the unknown auxin biosynthesis pathway (Shi et al., 2019; Zhou et al., 2022). Thus, it is suggested that these two genes were highly related to the gene function of this locus and further investigation was needed. A P-

10.3389/fpls.2023.1138988

TABLE 2 The information of the detected pod size SNPs in current study.

Trait	SNP	Year(s)	Chr	Pos	LOD	Marker R^2
PL	Pv_0026128	2020	1	19213953	3.63	22.32%
	Pv_0285901	2020	4	344060	3.71	23.90%
	Pv_0311899	2020	4	17510650	3.53	13.47%
	Pv_0429726	2021	5	14767260	3.93	16.58%
	Pv_0445220	2020 and 2021	5	25626390	4.16	20.33%
	Pv_0562916	2020	7	13389784	3.58	18.23%
	Pv_1016502	2021	11	1129181	3.56	21.11%
	Pv_1129860	2020	11	47823838	3.51	19.17%
	Pv_1154801	2020	11	51850889	3.60	22.37%
PT	Pv_0002282	2021	1	3846586	3.51	27.14%
	Pv_0029912	2020	1	21338661	3.94	21.39%
	Pv_0201428	2020	2	48802744	3.69	29.50%
	Pv_0212827	2021	3	4133573	3.51	22.87%
	Pv_0579990	2020	7	19691399	3.71	24.30%
	Pv_0684832	2020	8	8785254	3.57	19.69%
	Pv_0685388	2020	8	9098191	3.93	30.82%
PW	Pv_0008234	2020 and 2021	1	5648221	6.92	47.27%
	Pv_0127220	2020 and 2021	2	28035744	6.28	37.42%
	Pv_0193604	2020	2	43687336	6.60	38.01%
	Pv_0195014	2020	2	46751740	6.22	38.70%
	Pv_0203965	2020	2	48895245	6.23	35.07%
	Pv_0226301	2020 and 2021	3	12533721	6.14	48.33%
	Pv_0272703	2020 and 2021	3	32138995	6.04	51.31%
	Pv_0274424	2020 and 2021	3	33692550	6.83	43.40%
	Pv_0298015	2020 and 2021	4	8968938	6.00	44.63%
	Pv_0303974	2020 and 2021	4	13113489	6.02	39.84%
	Pv_0366735	2020 and 2021	4	46263029	7.06	45.80%
	Pv_0368967	2020	4	47052667	6.07	42.89%
	Pv_0421055	2020 and 2021	5	8368043	6.92	55.70%
	Pv_0428704	2020	5	14219936	6.37	46.13%
	Pv_0447805	2020 and 2021	5	27146151	6.05	44.38%
	Pv_0519622	2020 and 2021	6	18274170	6.42	34.04%
	Pv_0545768	2020	7	3001451	6.29	41.86%
	Pv_0547992	2020 and 2021	7	5917854	6.33	47.38%
	Pv_0602824	2020	7	25300279	6.70	51.51%
	Pv_0721251	2020 and 2021	8	31136317	6.09	36.65%
	Pv_0802255	2020 and 2021	9	7474435	6.48	48.03%
	Pv_1169490	2020 and 2021	11	53574929	6.05	49.57%
SPW	Pv_0015964	2020	1	12473517	3.65	17.25%

(Continued)

10.3389/fpls.2023.1138988

TABLE 2 Continued

Trait	SNP	Year(s)	Chr	Pos	LOD	Marker R^2
	Pv_0165916	2020	2	36567260	3.98	30.29%
	Pv_0223474	2020	3	11173316	3.68	27.31%
	Pv_0237496	2020	3	16882167	4.14	28.93%
	Pv_0527678	2021	6	26312586	3.53	24.43%
	Pv_0527777	2020	6	26833487	3.59	23.74%
	Pv_0686810	2020	8	9704115	3.52	27.79%
	Pv_0692879	2020	8	13885583	3.93	32.25%
	Pv_0791404	2020	8	63047615	3.73	19.24%
	Pv_0932953	2020	10	12316708	3.82	23.97%
	Pv_0933835	2020	10	12783478	3.91	31.57%
	Pv_0939702	2020	10	16283834	3.88	28.21%
	Pv_0940934	2020	10	16955046	4.24	31.88%
	Pv_0948680	2021	10	20652263	3.51	38.28%
	Pv_0959635	2020	10	27212413	4.06	35.19%
	Pv_0961930	2020	10	28382540	4.11	32.87%
	Pv_0968910	2020	10	32052969	3.93	32.43%
	Pv_0974939	2020	10	34687854	3.51	23.43%
	Pv_0977521	2021	10	35814018	3.52	17.54%

LOD, logarithm of the odds; PL, pod length; PT, pod thickness; PW, pod width; SPW, single pod weight.

TABLE 3 Candidate genes for the detected pod size SNPs.

Trait	SNP	Pos	Gene	start	end	Annotated function
PL	Pv_0285901	344060	Phvul.004G005400	315927	317696	cytochrome P450, family 96, subfamily A, polypeptide 10
			Phvul.004G005700	368531	370087	cytochrome P450, family 96, subfamily A, polypeptide 10
			Phvul.004G005800	372078	374079	cytochrome P450, family 96, subfamily A, polypeptide 10
			Phvul.004G005900	380970	382499	cytochrome P450, family 96, subfamily A, polypeptide 10
			Phvul.004G006000	386618	394940	cytochrome P450, family 96, subfamily A, polypeptide 10
			Phvul.004G006050	394161	395837	cytochrome P450, family 96, subfamily A, polypeptide 10
			Phvul.004G006100	401756	403445	cytochrome P450, family 96, subfamily A, polypeptide 10
			Phvul.004G006300	431954	433623	cytochrome P450, family 96, subfamily A, polypeptide 1
	Pv_0445220	25626390	Phvul.005G091200	25636647	25640316	galacturonosyltransferase 12
	Pv_1016502	1129181	Phvul.011G015300	1205296	1211056	P-loop containing nucleoside triphosphate hydrolases superfamily protein
PT	Pv_0684832	8785254	Phvul.008G088100	8689853	8700154	HEAT repeat; WD domain, G-beta repeat protein protein
	Pv_0685388	9098191	Phvul.008G090300	9078509	9081654	WRKY DNA-binding protein 33
PW	Pv_0008234	5648221	Phvul.001G052100	5695145	5699567	cytochrome P450, family 71, subfamily B, polypeptide 35
	Pv_0193604	43687336	Phvul.002G265400	43645971	43647134	WRKY DNA-binding protein 51
			Phvul.002G266400	43771606	43773653	WRKY DNA-binding protein 13
	Pv_0272703	32138995	Phvul.003G128800	32230627	32238415	auxin response factor 19

(Continued)

10.3389/fpls.2023.1138988

TABLE 3 Continued

Trait	SNP	Pos	Gene	start	end	Annotated function
	Pv_0303974	13113489	Phvul.004G076200	13119224	13119583	cytochrome P450, family 72, subfamily A, polypeptide 15
	Pv_0366735	46263029	Phvul.004G159300	46348261	46352994	cytochrome P450, family 71, subfamily B, polypeptide 34
			Phvul.004G159500	46359615	46361846	cytochrome P450, family 71, subfamily B, polypeptide 34
	Pv_0368967	47052667	Phvul.004G166000	46991847	46995082	RHO-related protein from plants 9
SPW	Pv_0223474	11173316	Phvul.003G072200	11117010	11145109	aminophospholipid ATPase 3
	Pv_0237496	16882167	Phvul.003G085400	16950726	16952504	cytochrome P450, family 71, subfamily B, polypeptide 34
	Pv_0527678	26312586	Phvul.006G157800	26263616	26264925	Adenosylmethionine decarboxylase family protein
			Phvul.006G159300	26336534	26342306	histidine-containing phosphotransmitter 1
	Pv_0686810	9704115	Phvul.008G094800	9681839	9684032	myb-like HTH transcriptional regulator family protein
	Pv_0961930	28382540	Phvul.010G075500	28283209	28293553	myb domain protein 1

SNP, single nucleotide polymorphism; PL, pod length; PT, pod thickness; PW, pod width; SPW, single pod weight.

loop containing nucleoside triphosphate hydrolases superfamily protein (*NOA1*) *Phvul.011G015300* was considered as a candidate gene of the *PL SNP Pv_1016502*, which is involved in growth regulation and hormonal signaling in plants (Zhao et al., 2015). The auxin response factor *Phvul.003G128800* is regarded as the candidate gene for the PW SNP Pv_0272703, whereas auxin response factor 18 (*ARF18*) affects the seed weight and silique length by regulating cell growth in polyploid rapeseed (Liu et al., 2015). In addition, we found that a total of 19 candidate gene models are involved in the biosynthesis and regulation of plant hormones, indicating that auxin-response pathways play key roles in pod size development in snap beans, which indicates clear pathways for the cloning of pod size genes in future.

In summary, we detected 57 genomic regions that were significantly associated with pod size in Chinese snap bean germplasms, which provides genomic resources for molecular breeding of pod size. The two developed KASP markers could be used directly in the marker-assisted selection of pod size traits. The results also enhanced our understanding of the molecular mechanism of pod development and established a springboard for future pod size gene cloning and its application in snap bean cultivation.

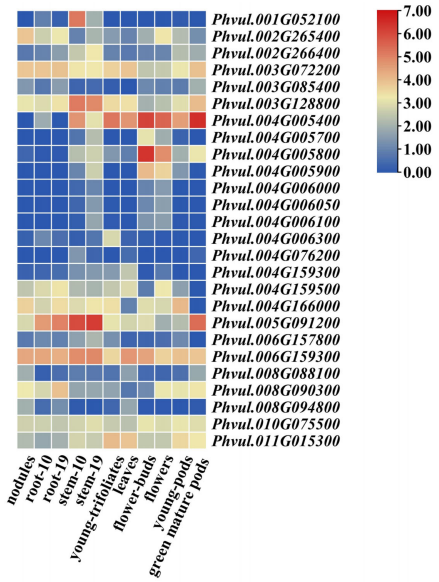


FIGURE 4
Expression patterns of 26 candidate genes in 11 tissues of common bean.

10.3389/fpls.2023.1138988

TABLE 4 The primer sequence information of the two KASP markers.

SNP	Primer_AlleleFAM	Primer_AlleleHEX	Primer_Common	Allele FAM	Allele HEX
Pv_0026128	GAAGGTGACCAAGTTCATG CTAGTTGCTTCAACCTGGAGC	GAAGGTCGGAGTCAACGGATT AGTTGCTTCAACCTGGAGT	TAAGTAATTGTTAT CGCTTGCTGACCT	C	T
Pv_0974939	GAAGGTGACCAAGTTCATGC TTCTGAGTCGTCTCGGGCACAA	GAAGGTCGGAGTCAACGGAT TTCTGAGTCGTCTCGGGCACAG	CTTGAGCGTCGGAGTGCCTTCT	A	G

KASP, kompetitive allele specific polymerase chain reaction; SNP, single nucleotide polymorphism.

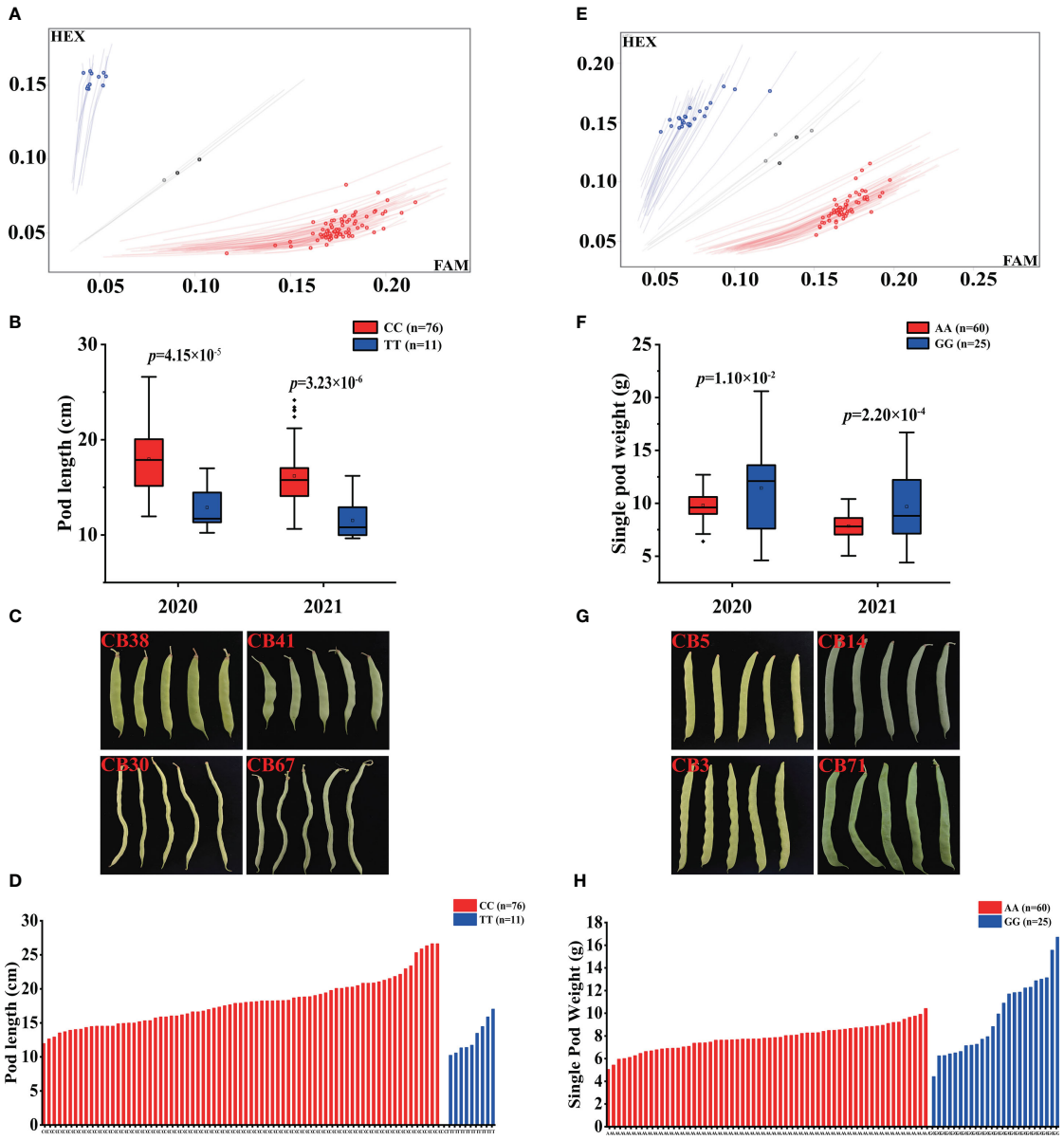


FIGURE 5
KASP markers development for PL SNP Pv_0026128 and SPW SNP Pv_974939. **(A)** genotype signal for KASP Pv_0026128 marker; **(B)** haplotypes analysis of Pv_0026128 in the diversity panel; **(C)** pods of representative accessions with different haplotypes on Pv_0026128; **(D)** the PLs and their genotype of each accession in the panel; **(E)** genotype signal for KASP Pv_974939 marker; **(F)** haplotypes analysis of Pv_974939 in the diversity panel; **(G)** pods of representative accessions with different haplotypes on Pv_974939; **(H)** the SPWs and their genotype of each accession in the panel. GWAS, genome-wide association study; KASP, kompetitive allele specific polymerase chain reaction; PL, pod length; SNP, single nucleotide polymorphism; SPW, single pod weight.

Data availability statement

All sequencing data of the 88 bean accessions are available from NCBI under BioProject accessions no. PRJNA934100.

Author contributions

XYW, GL, and JY conceived and designed the research; ML performed the experiments, analyzed the data, and wrote the draft manuscript; BW, XHW, and ZL contributed to germplasm collection; YW, JWa, YS, and WD contributed to phenotype evaluation. JD and JWu helped analyzing the data. XYW analyzed the data and revised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by the National Key R&D Program of China (Grant No. 2017YFE0114500) and the Third National Census and Collection of Crop Germplasm Resources in China (111821301354052030).

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1138988/full#supplementary-material

References

Baggett, J. R., and Kean, D. (1995). Inheritance of twisted pods in common bean (*Phaseolus vulgaris* l.). *J. Amer Soc. Hort Sci.* 120 (6), 900–901. doi: 10.21273/JASHS.120.6.900

Biswal, A. K., Hao, Z., Pattathil, S., Yang, X., Winkeler, K., Collins, C., et al. (2015). Downregulation of *GAUT12* in populus deltoides by RNA silencing results in reduced recalcitrance, increased growth and reduced xylan and pectin in a woody biofuel feedstock. *Biotechnol. Biofuels* 8, 41. doi: 10.1186/s13068-015-0218-y

Broughton, W. J., Hernandez, G., Blair, M., Beebe, M., Gepts, P., Vanderleyden, J., et al. (2003). Beans (Phaseolus spp.)-model food legumes. *Plant Soil* 252, 55–128. doi: 10.1023/A:1024146710611

Campa, A., Murube, E., and Ferreira, J. J. (2018). Genetic diversity, population structure, and linkage disequilibrium in a Spanish common bean diversity panel revealed through genotyping-by-sequencing. *Genes* 9 (11), 518. doi: 10.3390/genes9110518

Cui, X., Ge, C., Wang, R., Wang, H., Chen, W., Fu, Z., et al. (2010). The *BUD2* mutation affects plant architecture through altering cytokinin and auxin responses in arabidopsis. *Cell Res.* 20 (5), 576–586. doi: 10.1038/cr.2010.51

Delfini, J., Moda-Cirino, V., Dos Santos Neto, J., Ruas, P. M., Sant'Ana, G. C., Gepts, P., et al. (2021b). Population structure, genetic diversity and genomic selection signatures among a Brazilian common bean germplasm. *Sci. Rep.* 11 (1), 2964. doi: 10.1038/s41598-021-82437-4

Delfini, J., Moda-Cirino, V., Dos Santos Neto, J., Zeffa, D. M., Nogueira, A. F., Ribeiro, L. A. B., et al. (2021a). Genome-wide association study for grain mineral content in a Brazilian common bean diversity panel. *Theor. Appl. Genet.* 134 (9), 2795–2811. doi: 10.1007/s00122-021-03859-2

Dubos, C., Stracke, R., Grotewold, E., Weisshaar, B., Martin, C., and Lepiniec, L. (2010). MYB transcription factors in arabidopsis. *Trends Plant Sci.* 15 (10), 573–581. doi: 10.1016/j.tplants.2010.06.005

Feng, J., Wang, C., Chen, Q., Chen, H., Ren, B., Li, X., et al. (2013). S-nitrosylation of phosphotransfer proteins represses cytokinin signaling. *Nat. Commun.* 4, 1529. doi: 10.1038/ncomms2541

García-Fernández, C., Campa, A., Garzón, A. S., Miklas, P., and Ferreira, J. J. (2021). GWAS of pod morphological and color characters in common bean. *BMC Plant Biol.* 17, 21(1):184. doi: 10.1186/s12870-021-02967-x

González, A. M., Yuste-Lisbona, F. J., Saburido, S., Bretones, S., DeRon, A. M., and Lozano, R. (2016). Major contribution of flowering time and vegetative growth to plant production in common bean as deduced from a comparative genetic mapping. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.01940

Gupta, N., Zargar, S. M., Singh, R., Nazir, M., Mahajan, R., and Salgotra, R. K. (2020). Marker association study of yield attributing traits in common bean (*Phaseolus vulgaris* l.). *Mol. Biol. Rep.* 47 (9), 6769–6783. doi: 10.1007/s11033-020-05735-6

Hagerty, C. H., Cuesta-Marcos, A., Cregan, P., Song, Q., McClean, P., and Myers, J. R. (2016). Mapping snap bean pod and color traits, in a dry bean × snap bean recombinant inbred population. *J. Amer Soc. Hort Sci.* 141 (2), 131–138. doi: 10.21273/JASHS.141.2.131

Jing, H., Korasick, D. A., Emenecker, R. J., Morffy, N., Wilkinson, E. G., Powers, S. K., et al. (2022). Regulation of AUXIN RESPONSE FACTOR condensation and nucleo-cytoplasmic partitioning. *Nat. Commun.* 13 (1), 4015. doi: 10.1038/s41467-022-31628-2

Kamfwa, K., Cichy, K. A., and Kelly, J. D. (2015). Genome-wide association study of agronomic traits in common bean. *Plant Genome* 8, 1–12. doi: 10.3835/plantgenome2014.09.0059

Lamprecht, H. (1932). Beitrage zur genetik von phaseolus vulgaris. II. Über vererbung von hülsenfarbe und hülsenform. *Hereditas* 16, 295–340. doi: 10.1111/j.1601-5223.1932.tb02573.x

Lamprecht, H. (1947). The inheritance of the slender-type of phaseolus vulgaris and some other results. *Agri Hort Genet.* 5, 72–84.

Leakey, C. L. A. (1988). "Genotypic and phenotypic markers in common bean," in *Genetic resources of phaseolus beans*. Ed. P. Gepts (Dordrecht: Kluwer), 245–347.

Liu, J., Hua, W., Hu, Z., Yang, H., Zhang, L., Li, R., et al. (2015). Natural variation in *ARF18* gene simultaneously affects seed weight and silique length in polyploid rapeseed. *PNAS* 112 (37), E5123–E5132. doi: 10.1073/pnas.1502160112

Murube, E., Campa, A., Song, Q., McClean, P., and Ferreira, J. J. (2020). Toward validation of QTL associated with pod and seed size in common bean using two nested recombinant inbred line populations. *Mol. Breeding* 40 (1), 7. doi: 10.1007/s11032-019-1085-1

Myers, J. R., and Baggett, J. R. (1999). "Improvement of snap bean," in *Common bean improvement in the twenty-first century* (Switzerland: Springer: Basel), 289–329.

Nibau, C., Tao, L., Levasseur, K., Wu, H. M., and Cheung, A. Y. (2013). The *Arabidopsis* small GTPase AtRAC7/ROP9 is a modulator of auxin and abscisic acid signalling. *J. Exp. Bot.* 64 (11), 3425–3437. doi: 10.1093/jxb/ert179

Poulsen, L. R., López-Marqués, R. L., McDowell, S. C., Okkeri, J., Licht, D., Schulz, A., et al. (2008). The *Arabidopsis* P4-ATPase ALA3 localizes to the golgi and

requires a beta-subunit to function in lipid translocation and secretory vesicle formation. *Plant Cell.* 20 (3), 658–676. doi: 10.1105/tpc.107.054767

Powers, S. K., Holehouse, A. S., Korasick, D. A., Schreiber, K. H., Clark, N. M., Jing, H., et al. (2019). Nucleo-cytoplasmic partitioning of ARF proteins controls auxin responses in *Arabidopsis thaliana*. *Mol. Cell.* 76 (1), 177–190.e5. doi: 10.1016/j.molcel.2019.06.044

Prakken, R. (1934). Inheritance of colours and pod characters in *Phaseolus vulgaris* l. *Genetica* 16, 177–294. doi: 10.1007/BF02071498

Salem, M. A., Li, Y., Wiszniewski, A., and Giavalisco, P. (2017). Regulatory-associated protein of TOR (RAPTOR) alters the hormonal and metabolic composition of arabidopsis seeds, controlling seed morphology, viability and germination potential. *Plant J.* 92 (4), 525–545. doi: 10.1111/tpj.13667

Schmutz, J., McClean, P. E., Mamidi, S., Wu, G. A., Cannon, S. B., Grimwood, J., et al. (2014). A reference genome for common bean and genome-wide analysis of dual domestications. *Nat. Genet.* 46 (7), 707–713. doi: 10.1038/ng.3008

Shi, L., Song, J., Guo, C., Wang, B., Guan, Z., Yang, P., et al. (2019). A CACTA-like transposable element in the upstream region of *BnaA9.CYP78A9* acts as an enhancer to increase silique length and seed weight in rapeseed. *Plant J.* 98 (3), 524–539. doi: 10.1111/tpj.14236

Song, Q., Jia, G., Hyten, D. L., Jenkins, J., Hwang, E. Y., Schroeder, S. G., et al. (2015). SNP assay development for linkage map construction, anchoring whole-genome sequence, and other genetic and genomic applications in common bean. *G3 (Bethesda)* 5 (11), 2285–2290. doi: 10.1534/g3.115.020594

Tang, D., Gallusci, P., and Lang, Z. (2020). Fruit development and epigenetic modifications. *New Phytol.* 228 (3), 839–844. doi: 10.1111/nph.16724

Valdisser, P. A. M. R., Pereira, W. J., Almeida Filho, J. E., Müller, B. S. F., Coelho, G. R. C., de Menezes, I. P. P., et al. (2017). In-depth genome characterization of a Brazilian common bean core collection using DArTseq high-density SNP genotyping. *BMC Genomics* 18 (1), 423. doi: 10.1186/s12864-017-3805-4

Wallace, L., Arkwazee, H., Vining, K., and Myers, J. R. (2018). Genetic diversity within snap beans and their relation to dry beans. *Genes (Basel)* 9 (12), 587. doi: 10.3390/genes9120587

Wang, X., Niu, Y., and Zheng, Y. (2021). Multiple functions of MYB transcription factors in abiotic stress responses. *Int. J. Mol. Sci.* 22 (11), 6125. doi: 10.3390/ijms22116125

Wu, J., Wang, L., Fu, J., Chen, J., Wei, S., Zhang, S., et al. (2019). Resequencing of 683 common bean genotypes identifies yield component trait associations across a north-south cline. *Nat. Genet.* 52 (1), 118–125. doi: 10.1038/s41588-019-0546-0

Wu, X., Wang, B., Wu, S., Li, S., Zhang, Y., Wang, Y., et al. (2021). Development of a core set of single nucleotide polymorphism markers for genetic diversity analysis and cultivar fingerprinting in cowpea. *Legume Sci.* 3, e93. doi: 10.1002/leg3.93

Wu, X., Wang, B., Xin, Y., Wang, Y., Tian, S., Wang, J., et al. (2022). Unravelling the genetic architecture of rust resistance in the common bean *(Phaseolus vulgaris* l.) by combining QTL-seq and GWAS analysis. *Plants (Basel)* 11 (7), 953. doi: 10.3390/plants11070953

Xu, P., Moshelion, M., Wu, X., Halperin, O., Wang, B., Luo, J., et al. (2015). Natural variation and gene regulatory basis for the responses of asparagus beans to soil drought. *Front. Plant Sci.* 6. doi: 10.3389/fpls.2015.00891

Yuste-Lisbona, F. J., González, A. M., Capel, C., García-Alcázar, M., Capel, J., DeRon, A. M., et al. (2014). Genetic variation underlying pod size and color traits of common bean depends on quantitative trait loci with epistatic effects. *Mol. Breed.* 33 (4), 939–952. doi: 10.1007/s11032-013-0008-9

Zhang, X., Blair, M. W., and Wang, S. (2008). Genetic diversity of Chinese common bean *(Phaseolus vulgaris* l.) landraces assessed with simple sequence repeat markers. *Theor. Appl. Genet.* 117 (4), 629–640. doi: 10.1007/s00122-008-0807-2

Zhao, X., Wang, J., Yuan, J., Wang, X. L., Zhao, Q. P., Kong, P. T., et al. (2015). NITRIC OXIDE-ASSOCIATED PROTEIN1 *(AtNOA1)* is essential for salicylic acid-induced root waving in arabidopsis thaliana. *New Phytol.* 207 (1), 211–224. doi: 10.1111/nph.13327

Zhou, X., Zhang, H., Wang, P., Liu, Y., Zhang, X., Song, Y., et al. (2022). *BnaC7.ROT3*, The causal gene of *cqSL-C7*, mediates silique length by affecting cell elongation in *Brassica napus*. *J. Exp. Bot.* 73 (1), 154–167. doi: 10.1093/jxb/erab407

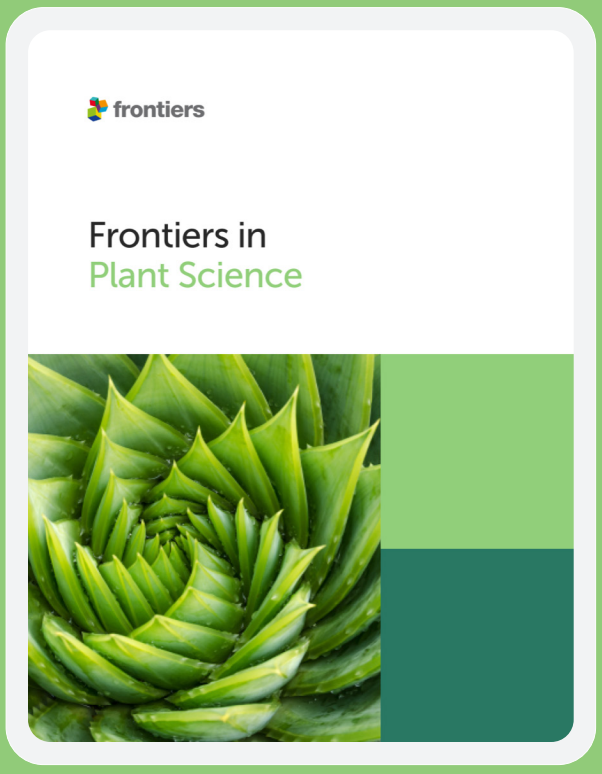
Frontiers in
Plant Science

Cultivates the science of plant biology and its applications

The most cited plant science journal, which advances our understanding of plant biology for sustainable food security, functional ecosystems and human health.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact