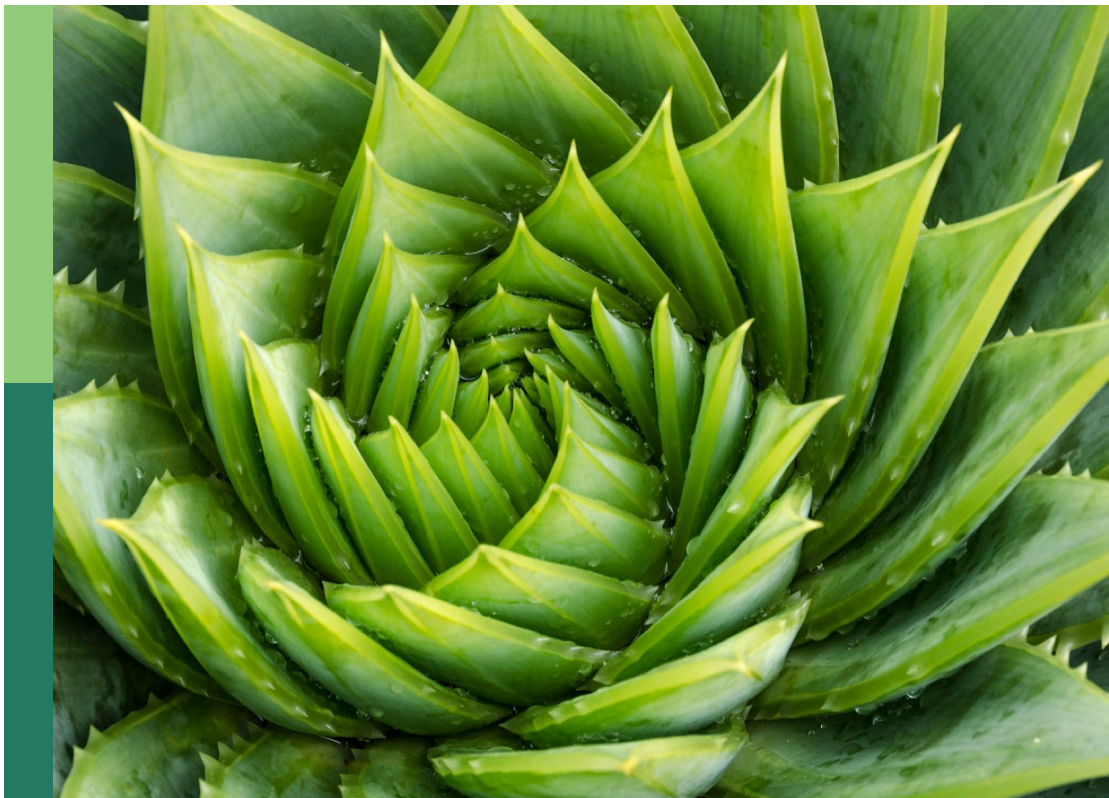
frontiers | Research Topics

Soybean breeding for abiotic stress tolerance: towards sustainable agriculture

Edited by
Huatao Chen, Donghe Xu, Henry T. Nguyen and Li Song

Coordinated by
Chengfu Su

Published in
Frontiers in Plant Science



ISSN 1664-8714
ISBN 978-2-8325-6315-1
DOI 10.3389/978-2-8325-6315-1

About Frontiers

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

Frontiers journal series

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

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Soybean breeding for abiotic stress tolerance: towards sustainable agriculture

Topic editors

Huatao Chen — Jiangsu Academy of Agricultural Sciences (JAAS), China

Donghe Xu — Japan International Research Center for Agricultural Sciences (JIRCAS), Japan

Henry T. Nguyen — University of Missouri, United States

Li Song — Yangzhou University, China

Topic coordinator

Chengfu Su — Qingdao Agricultural University, China

Citation

Chen, H., Xu, D., Nguyen, H. T., Song, L., Su, C., eds. (2025). *Soybean breeding for abiotic stress tolerance: towards sustainable agriculture*.

Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-6315-1

Table of contents

frontiers | Frontiers in Plant Science

TYPE Editorial
PUBLISHED 25 April 2025
DOI 10.3389/fpls.2025.1578039

OPEN ACCESS

EDITED AND REVIEWED BY
Jiedan Chen,
Chinese Academy of Agricultural Sciences, China

*CORRESPONDENCE
Huatao Chen
cht@jaas.ac.cn

RECEIVED 17 February 2025
ACCEPTED 10 April 2025
PUBLISHED 25 April 2025

CITATION
Chen H, Song L, Nguyen HT, Xu D and Su C (2025) Editorial: Soybean breeding for abiotic stress tolerance: towards sustainable agriculture.
Front. Plant Sci. 16:1578039.
doi: 10.3389/fpls.2025.1578039

Editorial: Soybean breeding for abiotic stress tolerance: towards sustainable agriculture

Huatao Chen[1]*, Li Song[2], Henry T. Nguyen[3], Donghe Xu[4] and Chengfu Su[5]

[1]Institute of Industrial Crops, Jiangsu Academy of Agricultural Sciences, Nanjing, China, [2]College of Horticulture, Yangzhou University, Yangzhou, China, [3]Division of Plant Sciences, University of Missouri, Columbia, SC, United States, [4]Biological Resources and Post-harvest Division, Japan International Research Center for Agricultural Sciences (JIRCAS), Tsukuba, Japan, [5]College of Agronomy, Qingdao Agricultural University, Qingdao, China

KEYWORDS

soybean, abiotic stress, tolerance, KASP, biotechnology

Editorial on the Research Topic
Soybean breeding for abiotic stress tolerance: towards sustainable agriculture

Soybean, with its high - protein and oil - rich seeds, is a vital agricultural commodity. However, the growing global demand, spurred by population growth and dietary changes, poses significant challenges in meeting production demand. Abiotic and biotic stresses, such as drought, salinity, extreme temperatures, and flooding, are major hurdles in soybean breeding.

To address these challenges, modern soybean breeding strategies aim to develop stress tolerant varieties. Advancements in molecular biology and biotechnology provide innovative approaches to this end. Techniques such as Kompetitive Allele-Specific PCR (KASP) molecular markers help breeders to accurately select stress - tolerant genes, while omics - based approaches provide insights into the molecular mechanisms underlying soybean's stress responses. These findings are presented in eleven papers on the Research Topic "Soybean Breeding for Abiotic Stress Tolerance: Towards Sustainable Agriculture."

Salt tolerance

Wang et al. investigate the genetic basis of salt tolerance in soybeans, a crucial aspect considering the global issue of soil salinization. A natural population of 283 soybean germplasms was used for this study. After identifying 180 mM NaCl as the optimal stress concentration, germination traits such as germination rate, energy, and index were measured under salt stress conditions. Through a genome-wide association study (GWAS), 1841 significant SNPs associated with these traits were identified, leading to the identification of 12 candidate genes. KASP markers were developed for specific SNPs. These findings offer valuable insights into soybean salt tolerance and offer genetic resources and a theoretical foundation for breeding salt-tolerant soybean varieties.

Xu et al. aimed to identify salt-tolerance-associated SNPs in soybean and evaluate genomic prediction for salt tolerance, using 563 germplasms from twenty-six countries. They identified four subpopulations (Q1-Q4) through relevant analyses. GWAS identified

10 SNPs on chromosomes 1, 2, 3, 7, and 16 that were significantly associated with salt tolerance, with eleven candidate genes located near 7 of these SNPs. Genomic prediction models including maBLUP, gBLUP, and sBLUP showed moderate - to - high r - values, and the GWAS-derived SNP marker set was effective for genomic selection. The genetic diversity analysis of salt-tolerant germplasms indicated a broad genetic background and highlighted the influence of geographic factors on salt tolerance in soybeans.

Drought tolerance

In response to the challenge of drought affecting soybean production in China, Jia et al. focused on a natural population of 264 Chinese soybean accessions. They treated these accessions with 15% PEG - 6000 during germination and employed a series of evaluation indexes, identified 17 drought - tolerant germplasms. Utilizing Genome - Wide Association Studies (GWAS), 92 SNPs and 9 candidate genes related to drought tolerance were discovered. Moreover, two KASP markers associated with drought tolerance were developed, which not only augment the soybean germplasm pool but also establish a crucial basis for molecular breeding of drought-resistant soybean varieties.

Flooding tolerance

Soybean is highly sensitive to flooding stress, and the Hypoxia Inducible Gene Domain (*HIGD*) gene family may play a role in plant responses to hypoxia. Geng et al. identified six *GmHIGD* genes in the soybean genome. The genes have conserved genomic structures and motifs. Chromosomal location and collinearity analysis revealed their distribution and potential evolutionary relationships. Cis - element analysis of promoters and TF identification suggested their involvement in growth, development, and stress responses. Expression analyses across various tissues and stresses (flooding, hypoxia, drought, salt) were conducted. Notably, GmHIGD3 was found to be localized in mitochondria, and its overexpression in *Arabidopsis* affected catalase activity and ROS content. Overall, this research provides valuable insights into the characteristics and potential functions of the *GmHIGD* gene family in soybean.

Shoot tolerance

Jia et al. investigate the impact of shade on soybean yield, with a focus on identifying shade-tolerant genomic loci and varieties. A natural population of 264 soybean accessions was subjected to a 30% light reduction treatment. GWAS was conducted on six agronomic traits and shade tolerance coefficients (STCs). Five high shade-tolerant germplasms were found, and a total of 733 significant SNPs associated with STCs of six traits were detected over two years. Four candidate genes related to shade tolerance were found. Additionally, KASP markers were developed for four SNPs, and haplotype analysis was performed. These results provide valuable genetic resources and new insights for soybean shade tolerance breeding and theoretical research.

Root architecture

In soybean, root architecture traits are vital for plant performance. The root length locus qRL16.1 on chromosome 16 has been previously reported. Through transcriptome analysis of near - isogenic lines (NILs), Kumawat et al. characterized two candidate genes, *Glyma.16g108500* and *Glyma.16g108700*, which exhibited higher expression in longer root accessions. The C-terminal domains of these genes are similar to those of C-terminally encoded peptides (CEPs) in *Arabidopsis*, known to regulate root length and nutrient response. Two polymorphisms located upstream of *Glyma.16g108500* were associated with root length traits. Synthetic peptide assays showed a positive effect of the predicted CEP variants on primary root length. These genes are specifically expressed in the root during the early growth stage and shown differential expression pattern only in the primary root. They hold potential for improving soybean to develop a strong root system under low moisture and nutrient conditions.

Seed traits

Soybean seed viability is crucial for both quality and production yet seed quality and germplasm preservation face challenges. Li et al. investigated the mechanisms of soybean seed aging by using aging-sensitive R31 and aging-tolerant R80 lines, subjecting them to artificial aging treatments of varying durations. Analyses of the transcriptome and metabolome revealed that the response to aging stress is associated with the phenylpropanoid metabolism pathway, in which caffeic acid plays a key role. Furthermore, soaking seeds in caffeic acid was found to enhance germination rates. These fundings provide a theoretical basis for future research on soybean seed aging mechanisms.

RNA modification

Despite the known significance of N6 - methyladenosine (m6A) RNA modification in regulating biological processes, its genome - wide identification and functional characterization in legumes like soybean have been lacking. Liu et al. used bioinformatics to identify thirteen m6A writer complex genes in soybean, which grouped into four families. They analyzed the characteristics, enzymatic activities, and expression patterns of these genes under abiotic stresses conditions, highlighting the roles of GmMTAs and GmMTBs in soybean's response to abiotic stress. This study establishes a foundation for further exploration of the functions of m6A modification in soybean.

Author contributions

HC: Writing – original draft, Writing – review & editing. LS: Writing – review & editing. HN: Writing – review & editing. DX: Writing – review & editing. CS: Writing – review & editing.

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.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 07 February 2024
DOI 10.3389/fpls.2024.1352465

OPEN ACCESS

EDITED BY
Zhenbin Hu,
Agricultural Research Service (USDA), United States

REVIEWED BY
Hengyou Zhang,
Chinese Academy of Sciences (CAS), China
Yuzhou Xu,
Kansas State University, United States

*CORRESPONDENCE
Huatao Chen
cht@jaas.ac.cn
Chengfu Su
chfsu2008@163.com

†These authors have contributed equally to this work

RECEIVED 08 December 2023
ACCEPTED 12 January 2024
PUBLISHED 07 February 2024

CITATION
Wang J, Zhou M, Zhang H, Liu X, Zhang W, Wang Q, Jia Q, Xu D, Chen H and Su C (2024) A genome-wide association analysis for salt tolerance during the soybean germination stage and development of KASP markers.
Front. Plant Sci. 15:1352465.
doi: 10.3389/fpls.2024.1352465

A genome-wide association analysis for salt tolerance during the soybean germination stage and development of KASP markers

Junyan Wang[1,2†], Miaomiao Zhou[1,2†], Hongmei Zhang[2], Xiaoqing Liu[2], Wei Zhang[2], Qiong Wang[2], Qianru Jia[2], Donghe Xu[3], Huatao Chen[2,4*] and Chengfu Su[1*]

[1]College of Agronomy, Qingdao Agricultural University, Qingdao, China, [2]Institute of Industrial Crops, Jiangsu Academy of Agricultural Sciences, Nanjing, China, [3]Japan International Research Center for Agricultural Sciences (JIRCAS), Tsukuba, Ibaraki, Japan, [4]Zhongshan Biological Breeding Laboratory (ZSBBL), Nanjing, China

Salt stress poses a significant challenge to crop productivity, and understanding the genetic basis of salt tolerance is paramount for breeding resilient soybean varieties. In this study, a soybean natural population was evaluated for salt tolerance during the germination stage, focusing on key germination traits, including germination rate (GR), germination energy (GE), and germination index (GI). It was seen that under salt stress, obvious inhibitions were found on these traits, with GR, GE, and GI diminishing by 32% to 54% when compared to normal conditions. These traits displayed a coefficient of variation (31.81% to 50.6%) and a substantial generalized heritability (63.87% to 86.48%). Through GWAS, a total of 1841 significant single-nucleotide polymorphisms (SNPs) were identified to be associated with these traits, distributed across chromosome 2, 5, 6, and 20. Leveraging these significant association loci, 12 candidate genes were identified to be associated with essential functions in coordinating cellular responses, regulating osmotic stress, mitigating oxidative stress, clearing reactive oxygen species (ROS), and facilitating heavy metal ion transport - all of which are pivotal for plant development and stress tolerance. To validate the candidate genes, quantitative real-time polymerase chain reaction (qRT-PCR) analysis was conducted, revealing three highly expressed genes *(Glyma.02G067700*, *Glyma.02G068900*, and *Glyma.02G070000)* that play pivotal roles in plant growth, development, and osmoregulation. In addition, based on these SNPs related with salt tolerance, KASP (Kompetitive Allele-Specific PCR)markers were successfully designed to genotype soybean accessions. These findings provide insight into the genetic base of soybean salt tolerance and candidate genes for enhancing soybean breeding programs in this study.

KEYWORDS

soybean, salt tolerance, germination stage, genome-wide association analysis, KASP marker

Introduction

Soil salinization constitutes a pressing global concern, posing a significant threat to crop growth and food production (Zhai et al., 2023). Recent statistics reveal that approximately 23% of cultivated land worldwide is affected by soil salinization, with 1.1 billion hectares of global land area afflicted by this issue. China is not immune to this challenge, with a saline soil area encompassing 36.9 million hectares, a substantial 10% of the global saline soil extent, and accounting for 5% of China's total available land (Zhao et al., 2022). China's saline soil total area is 36.9 million hm^2, accounting for 10% of the global saline soil, accounting for 5% of the country's available land area (Zhao et al., 2022). This issue manifests diversely across regions, including coastal saline soil and sea mud along the eastern coast, salt-affected soil in the Huang-Huai-Hai Plain, saline soil in the northeast plain, salt-impacted soil in the northwest inland areas, and desert salt soil in Qinghai and Xinjiang (Mao et al., 2020). In response to this critical concern, ensuring food security has prompted state initiatives aimed at optimizing available land resources through systematic planning of saline-alkali land, the selection of salt-tolerant crops for soil amelioration, and the preservation of cultivated land areas (Luan et al., 2023).

Soybean, a member of the legume family and the Papilionoideae stands as a pivotal cash crop, oilseed, and edible plant protein source in China. It also plays a crucial role as an industrial raw material (Li, 2011; Meng et al., 2021; Lu et al., 2023; Wang et al., 2023). Moreover, soybean holds a unique distinction as the cornerstone of Sino-U.S. agricultural trade relations, drawing significant attention from researchers. Because of saline land on soybean yield of serious damage to make our country's soybean production. Therefore, we cultivate salt-tolerant high yield soybean varieties of this work is particularly important.

Up to now, 1536 QTLs associated with salt tolerance have been reported, generally located on chromosomes 2, 3, 6, 8, 9, 12, 13, 14, and 17.(SoyBase.org) Two different materials were used to locate the QTL of soybean salt tolerance, and one major QT of salt tolerance was detected on chromosome 3 (Hamwieh et al., 2011). A total of 21 QTLs were identified, including 4 QTLs related to relative germination rate, 8 QTLs related to relative imbibition rate, and 9 QTLs related to relative germination index (Teng et al., 2022). Based on the analysis of 549 soybean materials, 11 ERF genes were upregulated, among which the $ERF158^{H1}$, $ERF166^{H2}$, and $ERF170^{H1}$ haplotypes were excellent allelic variants, which significantly promoted soybean salt tolerance(Gao et al., 2023). In this study, 257 soybean cultivars with 135 SSR markers were used to perform epistatic association mapping for salt tolerance.A total of 83 QTL-by-environment (QE) interactions for salt tolerance index were detected (Zhang et al., 2014). In the study, a population of 184 recombinant inbred lines (RILs) was utilized to map quantitative trait loci (QTLs) related to salt tolerance. A major QTL related to salt tolerance at the soybean germination stage named qST-8 was closely linked with the marker Sat_162 and detected on chromosome 8 (Yu et al., 2019).

Genome-wide association studies (GWAS) have proven to be a potent tool for investigating complex quantitative traits (Huang et al., 2012). The rapid advancements in modern molecular biology technology have further bolstered the application of molecular marker techniques in various crop breeding domains, including wheat (Kun Dziayi Turhan et al., 2021), rice (He et al., 2023), maize (Ma et al., 2023), sorghum (Gao et al., 2023), rapeseed (Xiao et al., 2023), and other molecular breeding arenas. Soybean exhibits a multitude of complex quantitative traits, often under the control of multiple genes and influenced by both genotype and environmental factors (Kun Dziayi Turhan et al., 2021). Notably, Liang Tengyue et al. (Liang et al., 2023) conducted a genome-wide association analysis on 395 soybean germplasm resources using the GAPIT tool, identifying nine SNPs closely linked to single plant grain weight under low phosphorus conditions. Yang Hao et al. (Yang et al., 2023) conducted a study in the Sichuan-Chongqing region employing 135 SSR markers and 107,081 effective SNP markers for genotyping from 227 soybean varieties and detected 51 and 70 site significantly associated with fertility traits through comprehensive whole-genome association analysis. Compared with the seedling stage, research on the correlation analysis of salt tolerance at the germination stage of soybean has just begun.

The KASP (kompetitive allele-specific PCR) molecular marker is a new SNP typing method based on allele-specific amplification and high-sensitivity fluorescence detection. KASP is characterized by low cost and high throughput, and the accurate double-allele genotyping of SNP and InDel sites through specific matching of primer terminal bases. The method is widely used in molecular marker-assisted selection of rice, wheat, soybean, and other crops (Ertiro et al., 2015; Tian et al., 2018; Cheon et al., 2020; Jiang et al., 2021).

In this study, 283 soybean germplasm resources were used as materials. Under simulated NaCl salt stress, the germination rate, germination energy, and germination index at the germination stage were used as the screening parameters. The integration of genome-wide association analysis (GWAS) allowed us to identify pivotal site associated with soybean salt tolerance in a high-throughput manner. We further harnessed this knowledge to develop KASP markers, leveraging the significant association SNPs to facilitate early selection in the quest for salt-tolerant soybean breeding. This approach significantly alleviates the workload associated with soybean breeding efforts, expediting progress and advancing the field of soybean salt-tolerant molecular breeding. Our findings represent a valuable reference for future research on soybean salt tolerance and the selection and breeding of novel, salt-tolerant soybean varieties.

Materials and methods

A natural soybean population containingused 283 representative soybean germplasms, including 52 landraces, 212 cultivars, and 19 wild soybeans, was used in this study. To identify an optimal stress concentration for evaluating salt tolerance in soybean germination, eight randomly selected varieties from our study's test materials underwent a preliminary germination assay. Each variety was subjected to three replicate tests, with a concentration gradient spanning 0, 30, 60, 90, 120, 150, and 180 mM NaCl. Our test results revealed that at a concentration of 150

mmol/L NaCl, the germination rate and other measured parameters exhibited noticeable inhibition. Further increase of the NaCl concentration to 180 mM/L elicited significant disparities in the germination rate, germination energy, and germination index of the materials. Consequently, we identified 180 mmol/L NaCl as the optimal stress concentration for our subsequent experiments.

The germination test was carried out in a dedicated germination room, using a 3×4 grid layout of small squares, with each grid covered by 25 g vermiculite. 50 healthy, full, and pest-free seeds of the same size were used for each condition. 90 ml 0mM and 180 mM Nacl solutions were used for the stress treatment. The seeds were spread on the vermiculite and watered with the treatment solution then covered with 3-4 layers of filter paper soaked with treatment solution. The count of germination seeds was recorded every 24 hours for 7-8 days. This protocol was conducted with three biological replicates per material. Utilizing an established formula, the relative salt damage rate for each germination parameter, including germination rate, germination energy, and germination index, was calculated.

Several key parameters were calculated to assess soybean germination under salt stress conditions, providing insights into salt tolerance:

Germination Rate (GR): The germination rate, expressed as a percentage, was calculated using the formula: GR (%) = (Nt/N) × 100, where Nt represents the number of seeds germinated per grid on day t, and N represents the total number of seeds per grid for testing (unit: %) (Wang et al., 2019).

Germination Index (GI): The germination index was determined using the formula: GI = ∑Gt/Dt, where Gt represents the number of seeds germinated per grid on day t, and Dt signifies the number of days in the germination test up to day (Wang et al., 2019).

Germination Energy (GE): Germination energy, also expressed as a percentage, was calculated as GE (%) = N3/N × 100, where N3 represents the number of seeds germinated per grid on the 3rd day, and N represents the total number of seeds per grid for testing (unit: %) (Wang et al., 2019).

Relative Salt Damage Index (ST): The relative salt damage index was derived using the formula: ST = S/C. Here, C represents the control germination rate, germination index, and germination energy, while S signifies the germination rate, germination index, and germination energy under salt treatment (Wang et al., 2019).

Generalized heritability $h^2 = \sigma_g^2 / \left(\sigma_g^2 + \frac{\sigma_{ge}^2}{n} + \sigma^2 / nr \right)$, where σ_g^2 $(g = 1, 2, 3 \ldots 264)$ is the genotype variance of the test material, σ_{ge}^2 $(e = 1, 2)$ is the variance of the interaction between the genotype and the environment of the test material, σ^2 is the error variance, n is the number of environments, and r is the number of replicates (Yuan et al., 2023).

Genome-wide association analysis

We resequenced 283 materials in the early stage, achieving an average sequencing depth of 12.4 ×, and yielding a high-density physical map containing a total of 2,597,425 SNPs (Zhang et al., 2021). Genome-wide association analysis was calculated using the GAPIT algorithm package in R (Lipka et al., 2012), and the general linear model (GLM) (Liu et al., 2008) was used for genome-wide association analysis SNPs with -LogP values ≥ 5 are considered to be significant association sites.

Haplotype and candidate gene analysis

We delineated chromosome intervals based on the target genes, thereby generating a dedicated SNP annotation file and corresponding genotype data. The target interval was methodically classified into five distinct categories, namely the gene-related region (including exons, stopgain, stoploss, splicing, etc.), intronic and UTR regions, upstream and downstream flanking sequences, and intergenic regions. Haplotype analysis was subsequently conducted on these different types of SNP sites. To construct haplotype networks, we employed PopARTv1.7 software. The entire haplotype analysis process was carried out using the R programming language.

After the SNPs significantly associated with salt tolerance traits in soybean germination were identified, we referenced soybean genome information in the online database Phytozome13 (https://phytozome-next.jgi.doe.gov/info/Gmax_Wm82_a2_v1). The genes related to the control of soybean plant height within 120 kb of the SNPs were identified For the analysis of specific population structure, see the research report of our laboratory (Zhang et al., 2021), and the candidate genes were identified by Blastp comparison with the gene sequences in the Arabidopsis genome database.

Development of KASP markers

Using the Primer-BLAST function of NCBI (https://www.ncbi.nlm.nih.gov/), KASP-PCR amplification primers were designed based on the SNP site S05_41921861 (A/C) and S02_6088007 (A/G) significantly associated with the germination rate, germination energy, germination index of soybean. Each pair of primers consists of two specific forward primers F1 and F2 and a generic reverse primer R. F1 and F2 contain 6-carboxyfluorescein (FAM) and hexachloro-6-methylfluorescein (HEX) fluorescent linker sequences (underlined), respectively. The primer sequences were synthesized by Qingke Biotech (Nanjing).

PCR procedure used for genotyping with KASP markers

Amplification PCR was performed using KASP V4.0 2×Mastermix (LGC, England), The program is: predenaturation at 94 °C for 15 minutes; Denaturation at 94 °C for 20 s, extension at 61-55 °C for 1 minute, with a decrease of 0.6 °C per cycle for 10 cycles; Denaturation at 94 °C for 20 s, extension at 55 °C for 1 minute, 26 cycles.

RNA extraction and reverse transcription

Take 0.1g soybean seedlings were quickly ground into powder in liquid nitrogen, and RNA was extracted by Trizol method, Using RNA as template, cDNA was synthesized by HiScript II 1st Strand cDNA Synthesis Kit kit (Vazyme Biotech).

Determination of candidate gene expression levels

Two salt-tolerant materials were selected to detect gene expression levels, we employed quantitative reverse transcription PCR (RT-PCR) reactions using the Lis system. The internal reference gene selected for this analysis was Tubulin (GenBank: KRG91143.1). The primer sequences for both the candidate gene and the internal reference gene can be found in Table 1.

For qRT-PCR, a Gentier96E fluorescence quantitative PCR instrument (purchased from Xi'an Tianlong Co., LTD.) was used for the amplification reaction. The reaction system was prepared as follows: 25 μL 2×Phanta Max Buffer, 1 μL dNTP Mix (10mM each), 3 μL cDNA, 2 μL forward primer, 2 μL reverse primer, 1 μL Phanta Max Super-Fidelity DNA Polymerase (1U/μL), 16 μL water. The amplification detection was conducted in 96-well plates. The reaction procedure was as follows: denaturation at 95°C 10 s; denaturation at 56°C for 20 s, annealing at 72°C for 20 s for a total of 40 cycles.

The expression levels of the target gene were assessed by comparing CT (Cycle threshold) values. When the primers for the target gene of interest exhibited similar amplification coefficients to those of the internal reference gene, the relative expression of the target gene in each sample was calculated using the formula $2^{-\Delta\Delta CT}$. Here, ΔΔCT was determined as $(C_{T,\ Target} - C_{T,\ Tubulin})_{genotype} - (C_{T,\ Target} - C_{T,\ Tubulin})$calibrator, allowing for precise evaluation of gene expression levels.

Results

Statistical analysis of soybean germination phenotype

This study encompassed the investigation of 283 soybean materials with a focus on salt tolerance traits during the germination period. Three key germination-related traits, namely Germination Rate (GR), Germination Energy (GE), and Germination Index (GI) were recorded. Statistical analysis was conducted on these fundamental indexes of the materials, leading to the derivation of relative indexes, specifically Relative Germination Rate (RGR), Relative Germination energy (RGE), and Relative Germination Index (RGI). The findings are detailed in Table 2. Our analysis revealed substantial phenotypic variation in the relative germination traits among the 283 soybean materials, in both 2022 and 2023. Over these two years, RGR, RGE, and RGI exhibited ranges of 0.05 to 1.00, 0.00 to 1.00, and 0.04 to 1.00, respectively. The coefficient of variation (CV) for these traits ranged from 31.81% to 50.60%, and the generalized heritability (h^2) ranged from 63.87% to 86.48%, highlighting the impact of interactions between plants, environmental factors, and the interplay between plants and the environment on these traits. The observed generalized heritability for each phenotype underscores the quantitative nature of these traits, which are governed by multiple genes. Additionally, it signifies genuine genetic differences in the reproductive period within the population, making them conducive for further analysis, particularly in the context of association studies.

Box line plot and frequency distribution analysis of salt tolerance traits during the germination stage

The box line plot for the relative germination rate, relative germination energy, and relative germination index under two years

TABLE 1 Sequences of specific primers for qRT-PCR.

serial number	Gene ID	F	R
1	*Glyma.02G067600*	ACAGCATGGGGAGGAAGGTA	CGGAGGAGTGTCCGGATAGA
2	*Glyma.02G067700*	GCGAGTTTGTCCGAGACCAT	TAGCCGTCCCTCCATCGAAT
3	*Glyma.02G068300*	CGATGCACCCAATGATGCTG	TAGGTGGTGGAGACGACGAT
4	*Glyma.02G068700*	TCACAAGGTCGGAAAGCGAG	GTACTGCAACTGCACAAGGC
5	*Glyma.02G068900*	ATGTGCCTACTTGGGCCTTT	CCCGGTTCTGTTTCCCAAGA
6	*Glyma.02G069400*	CCTTGCTGAGCTGCTTTTGG	CTCCTCTTCCAGCTTCCGAC
7	*Glyma.02G070000*	CCAACCTCTTGGATGCCACA	TCCATGTTTGAAAGGTGGCG
8	*Glyma.05G244600*	AGAGAGCGAGTTTGTGCTCC	GCTGGCACTCTTCAACAAGC
9	*Glyma.05G245000*	TGGCTGGTGATCATTGGACC	ATTGATCGTGGCAACGGGAT
10	*Glyma.05G246400*	TGCGTCGTTAAGATGGGCAA	CCCACTGGGGAGGTCTTCTA
11	*Glyma.09G044300*	TGAAAGCGAGCAAGCGAAAC	TGCACTCCTTCAAGGCCAAA
12	*Glyma.09G045200*	CAAGAGCAGCAACAACTCGC	CATTCACCTGGCCCACAAGA

10.3389/fpls.2024.1352465

TABLE 2 Descriptive statistics of three germination-related traits in soybean populations under NaCl conditions.

Year	Trait	Max	Min	Mean	SD	CV(%)	h^2(%)
2022	RGR	1.00	0.05	0.63	0.22	34.90	84.78
	RGE	1.00	0.02	0.45	0.23	50.60	
	RGI	1.00	0.04	0.44	0.18	40.17	63.87
2023	RGR	1.00	0.09	0.64	0.20	31.81	
	RGE	1.00	0.00	0.47	0.24	49.95	86.48
	RGI	1.00	0.04	0.48	0.20	45.60	

84.78, 63.87, 86.48 are the respective value of RGR, RGE and RGI in 2022 and 2023 correspond to generalized heritability.

of salt stress treatment (as shown in Figure 1) revealed no disparities in the relative amplitudes for each index between the two years. These calculations for the 283 soybean germplasms were executed using Microsoft Excel 2016. Furthermore, we plotted frequency distribution and density curves (illustrated in Figure 2) for the relative germination rate, relative germination energy, and relative germination index. Evident from these representations is the varying extent of suppression in several indicators under salt stress. The histograms displaying the phenotypic data exhibit characteristics resembling approximately normal distribution.

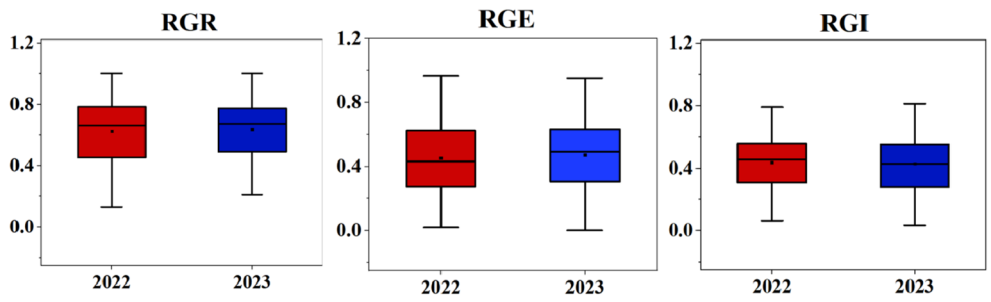


FIGURE 1
Differences in RGR, RGE, and RGI of soybean germplasm two years. Asterisks indicat significant differences compared with corresponding control(*P < 0.05, **P < 0.01,***P < 0.001).

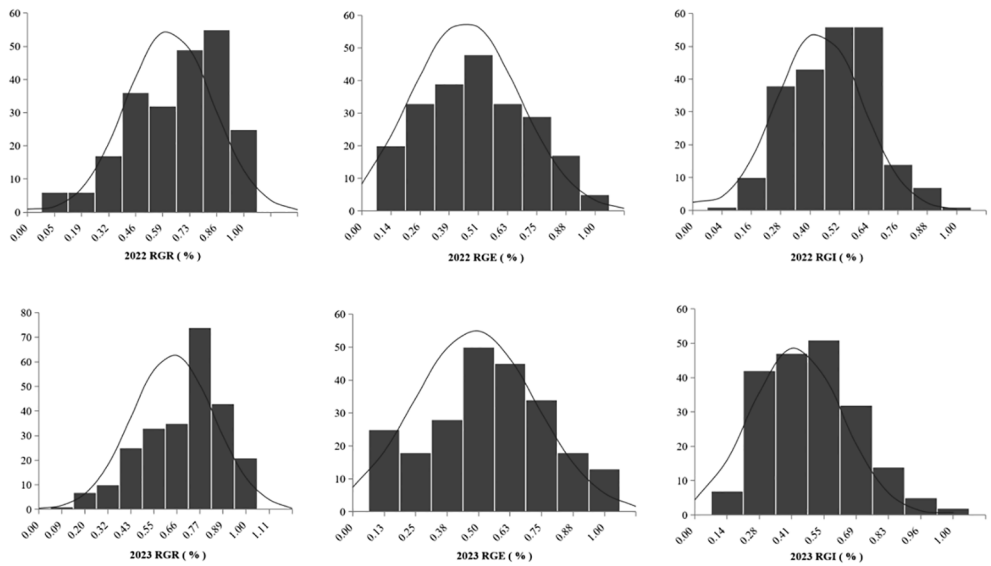


FIGURE 2
Frequency distribution of RGR, RGE, and RGI in soybean germination.

This implies that the natural population of soybeans within our study material has rich genetic variation, making it well-suited for subsequent genome-wide association analyses.

GWAS analysis of salt-tolerance-related traits in soybean germination

Within this study, we integrated the phenotypic results encompassing germination rate, germination energy, germination index, and their corresponding relative indexes (RGR, RGE, RGI) with sequencing data. We used a Generalized Linear Model (GLM) (Liu et al., 2008) to conduct genome-wide association analysis through the GAPIT package in R and created Manhattan plots and QQ plots representing the associated indicators (Figure 3). In 2022, a total of 447 SNPs (−log10P>5) closely associated with the soybean germination stage were detected, of which 269 SNP sites were associated with relative germination energy and mainly distributed on chromosomes 2, 5, and 20. Conversely, the sites least associated with the relative germination index were mainly distributed on chromosomes 2, 6, and 20. In 2023, a total of 1841 SNPs closely associated with soybean germination were detected. SNP site was most associated with relative germination rate, with 1512, and mostly distributed on chromosome 5. The sites associated with the relative germination index were mainly distributed on chromosomes 9 and 20. SNPs with significant correlation among traits during germination are detailed in Table 3.

Haplotype and candidate gene analysis

To study the phenotypic impact of allele variations at the most significantly associated SNP sites, haplotype analysis was conducted for these high-threshold SNP sites associated with salt tolerance traits during the germination stage in both 2022 and 2023. It was observed that at the SNP site S05_41921861, the allele variation was A/C, with the average relative germination rate for S05_41921861-A measuring 0.52, a significant reduction compared to 0.71 for S05_41921861-C. Meanwhile, at SNP site S02_6088007, allele variation was A/G, and the average relative germination energy for S02_6088007-A was notably lower at 0.42, as opposed to 0.66 for S02_6088007-G. Lastly, for SNP site S09_3907313, the allele variation was T/C, and the average relative germination index for S09_3907313-C was 0.40, again demonstrating a decrease compared to S09_3907313-T, which exhibited an average of 0.53. This trend in allele variation and its phenotypic effects remained consistent between the years 2022 and 2023 (Figure 4).

Candidate gene screening and function prediction were performed in the range of 120 kb upstream and downstream of SNP sites significantly associated with soybean germination tolerance (-log10(P)≥5). Concerning the gene function annotation information of the soybean genome, 12 candidate genes significantly associated with salt tolerance in soybean germination were identified (Table 4). These candidate genes were found to be involved in a wide range of critical functions, including coordinating cellular responses, regulating osmotic stress, mitigating oxidative stress, facilitating the clearance of reactive oxygen species (ROS), and functioning as heavy metal ion transporters. Collectively, these genes play pivotal roles in promoting plant development, enhancing stress tolerance, and ensuring normal growth and development of plants.

Development of KASP markers

KASP markers were developed for SNP sites S05_41921861 (A/C) and S02_6088007 (A/G), which exhibited significant associations with salt tolerance in soybean germination (Table 5). The variation at the S05_41921861 (A/C) site corresponds to the Glyma.05G244600 gene, and its annotation reveals a role in coordinating cellular responses facilitating normal plant growth and development, immune responses, and the capacity to respond to stress. Similarly, the variation at the S02_6088007 (A/G) site is associated with the gene Glyma.02G067600, and its annotation indicates involvement in the upregulation of stress responses. Under salt stress conditions, it triggers the expression the expression of GAOX20, encoding adC7-GA inhibitors. The designed KASP marking series is shown in Table 5. The genomic DNA from the selected soybean germplasm was extracted, and the KASP primers, designed for the aforementioned SNP sites, were employed in PCR reactions utilizing genomic DNA as the reaction template. After the completion of the reaction, fluorescence data results were directly read on the real-time PCR system. The genotyping of 48 selected soybean germplasms was executed using the KASP-labeled primers designed for the S05_41921861 (A/C) and S02_6088007 (A/G) sites. The results, illustrated in Figure 5, demonstrate the separation of the two different genotypes by PCR.

Determination of candidate gene expression levels

The reverse transcription PCR (Figure 6) and the rich soybean genome information enabled us to identify *Glyma.02G067700*, Glyma.02G068900, and *Glyma.02G070000* as the genes associated with salt tolerance in soybeans within this population. Incorporating comparative genomics studies of these three candidate genes with other crops and model plants, we uncovered the following insights:

Glyma.02G067700 codes for a MYB family protein, indicating its role as a key factor in regulatory networks governing development, metabolism, and responses to biotic and abiotic stresses (Shao et al., 2020).

Glyma.02G068900 encodes xyloglucan endo-transglycosylase/hydrolase (Ph XET/H), which regulates seed germination by facilitating the accumulation of Ph XET protein via GA-mediated pathways. This gene plays a pivotal role in endosperm weakening and embryonic expansion during seed germination, falling within the glycosyl hydrolase family 16 (Jacqueline et al., 1993).

10.3389/fpls.2024.1352465

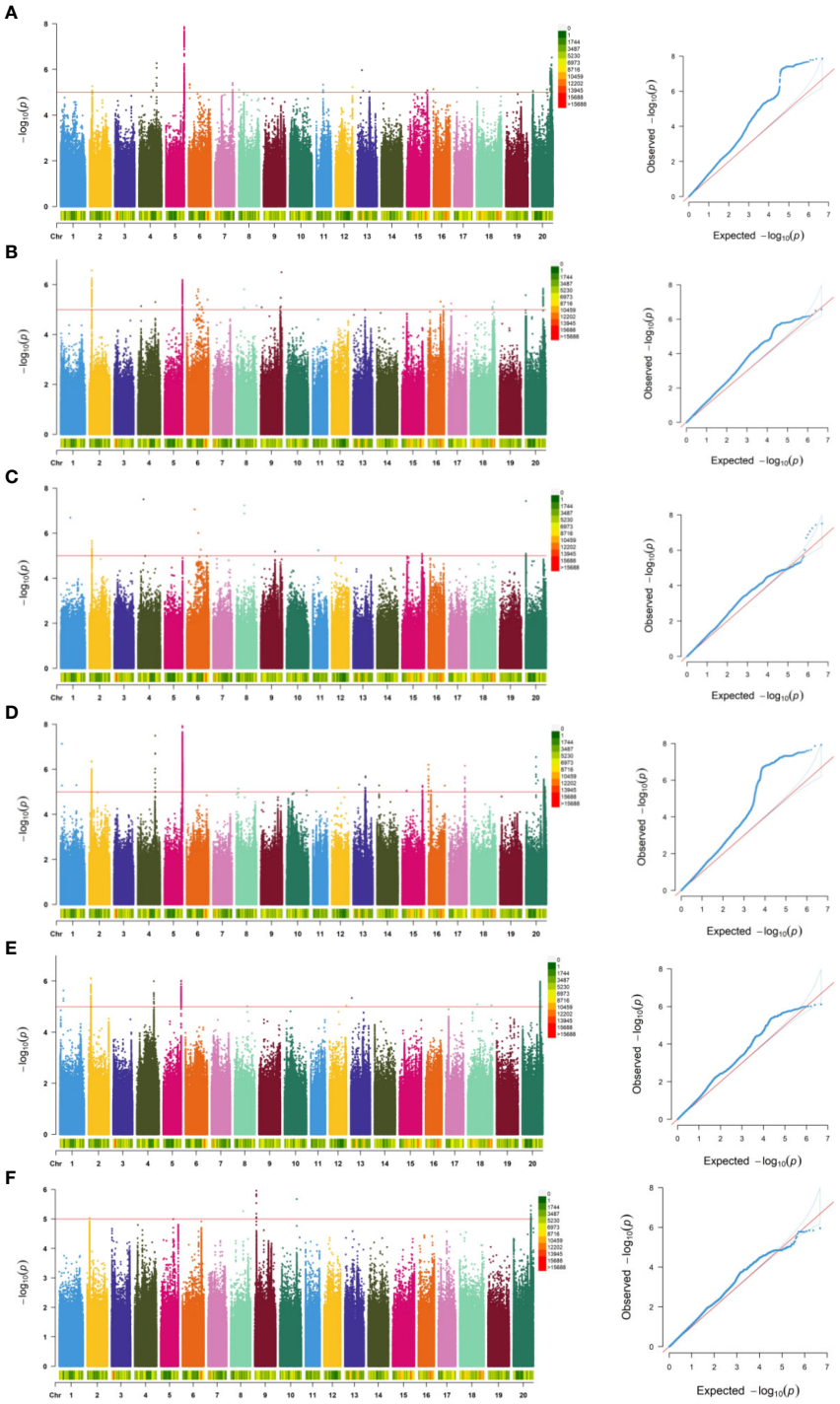


FIGURE 3
Genome-wide association analysis results of RGR, RGE, and RGI in the natural population of soybean over two years. The solid red lines in the Manhattan plots represent the significant threshold -log10(P)=5.0. **(A-F)** respectively is 2022RGR, 2022RGE, 2022RGI, 2023RGR, 2023RGE, 2023RGI.

Glyma.02G070000 codes for an NAC transcription factor, a plant-specific family of transcription factors known for their essential roles in various biological processes (Yuan et al., 2019).

These three genes play an important role in responding to biotic or abiotic stresses, as well as in regulating plant growth and development, and osmoregulation.

Discussion and conclusion

The research and development of salt-tolerant soybeans for saline-alkali soybean production and the expansion of planting areas through various strategies are pivotal steps in addressing the issue of insufficient soybean production capacity in China. These

10.3389/fpls.2024.1352465

TABLE 3 Statistics of GWAS analysis results of germination correlation traits.

Year	Trait	Chromosome	Position interval	Position interval	Peak SNP position	(-log Pmax)	PVE (%)
2022	RGR	4	6	40079022-40641706	40641706	6.27	12.42
		5	137	41755110-42233137	41921861	7.87	16.9
		20	1	2161697-45389200	45231603	6.53	10.67
	RGE	2	82	5966798-6088986	6088007	6.57	11.53
		5	143	41782306-41954581	41912280	6.19	10.76
		20	44	981655-41811347	41792703	5.84	10.06
	RGI	2	27	6062303-6088007	6073517	5.66	9.95
		6	3	18099861-32344653	26921222	7.06	10.66
		20	4	981655-989432	981655	7.43	13.68
2023	RGR	2	116	5966798-6257205	6086046	6.36	10.79
		5	1335	41763734-42233476	41921861	7.92	13.94
		20	61	23510283-44725190	23510343	6.55	10.32
	RGE	2	84	5966798-6257205	6086076	6.11	10.63
		5	133	41871375-41954581	41937985	6.01	10.43
		20	87	41496276-41927972	41792703	5.99	10.37
	RGI	9	13	3852644-3908645	3907313	5.96	12.14
		20	12	41656423-41927972	41759271	5.3	10.92

efforts bear significance for China's food security. Several studies have contributed to our understanding of salt tolerance mechanisms during soybean germination, shedding light on the physiological changes that occur under salt stress. Hao Xuefeng et al. examined salt tolerance and the salt tolerance mechanism in soybean seeds during germination and found changes in parameters such as SOD, POD, and MDA in the radicle germ in response to increasing NaCl concentrations. This research underscores the

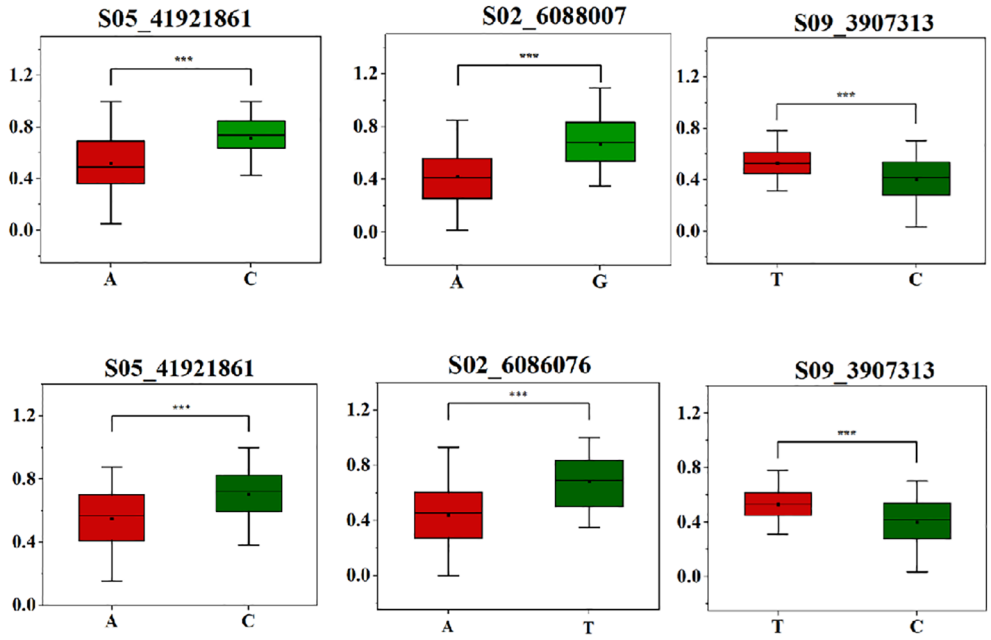


FIGURE 4
SNP haplotype analysis associated with salt-tolerant traits in a soybean natural population. A is a significant haplotype in 2022, B is a significant haplotype in 2023.Asterisks indicat significant differences compared with corresponding control X axis is the different haplotypes of each point, and Y axis is RGR, RGE and RGI respectively (*P< 0.05, **P< 0.01, ***P< 0.001).

TABLE 4 Functional annotations of candidate genes related to salt tolerance in soybean germination.

Gene ID	Homologs	Functional annotation
Glyma.02G067600	*AT5G13910.1*	Integrase-type DNA-binding superfamily protein
Glyma.02G067700	*AT3G02050.1*	K^+ uptake transporter 3
Glyma.02G068300	*AT3G02065.3*	P-loop containing nucleoside triphosphate hydrolases superfamily protein
Glyma.02G068700	*AT5G27690.1*	Heavy metal transport/detoxification superfamily protein
Glyma.02G068900	*AT5G13870.1*	xyloglucan endotransglucosylase/ hydrolase 5
Glyma.02G069400	*AT4G18710.2*	Protein kinase superfamily protein
Glyma.02G070000	*AT3G04070.2*	NAC domain-containing protein 47
Glyma.05G244600	*AT3G18040.1*	MAP kinase 9
Glyma.05G245000	*AT1G18180.1*	Protein of unknown function (DUF1295)
Glyma.05G246400	*AT3G17980.1*	Calcium-dependent lipid-binding (CaLB domain) family protein
Glyma.09G044300	*AT3G16630.2*	P-loop containing nucleoside triphosphate hydrolases superfamily protein
Glyma.09G045200	*AT1G56210.1*	Heavy metal transport/detoxification superfamily protein

existence of specific salt tolerance mechanisms and associated physiological changes during germination (Hao et al., 2013).

Some studies have identified the impact of salt stress on organelle formation, including chloroplasts and endoplasmic reticulum, resulting in varying degrees of influence on growth traits such as root length, hypocotyl length, and lateral root numbers. Ultimately, this process inhibited soybean germination (Liao et al., 2013). demonstrated that high-concentration NaCl stress significantly impeded water absorption in soybean seeds, leading to reduced amylase and protease activity, further elucidating the complexities of salt stress on germination (Xu et al., 2017). Kan found that 22 SSR markers and 11 related QTL sites were closely linked to salt tolerance in soybean germination, and localized on chromosomes 2, 7, 8, 10, 17, and 18 (Kan et al., 2016). research identified a total of 31 salt-tolerant-related QTLs through linkage analysis of ST-IR, ST-GI, ST-GE and ST-GR during the germination stage of an NJIKY population, mainly distributed on chromosomes 1, 2, 7, 8, 10, 15, 17 and 18 (Zhang et al., 2018). Furthermore, Kan used a natural population of 191 local soybean varieties and 1356 SNP markers to perform genome-wide association analysis. Their work identified five candidate genes closely linked to salt tolerance during soybean germination (Kan et al., 2015).

In our pursuit of identifying outstanding salt-tolerant soybean varieties and enhancing soybean yield, this study conducted a comprehensive analysis of germination traits, including germination rate, germination energy, germination index, and their relative values, across a diverse set of 283 soybean germplasm resources subjected to salt stress at the germination stage. Our investigation revealed that the germination traits within this population exhibited a rich and continuous distribution. At the same time, using the high-density SNP physical map combined with phenotype and genotype data for genome-wide association analysis, a total of 1841 SNP sites significantly associated with soybean germination stage were detected on chromosomes 2, 5, 6, 9, and 20. Notably, the loci located on chromosome 5 were repeatedly detected in 2 environments, and the genetic variation explainable by the GWAS signal reached 14.00%, marking it as a prominent genetic locus. MAP kinase 9 may be an effector gene for this site.In the same chromosomal interval as the results of other researchers, there may be allelic variation of the same QTL.Chromosomes 2 and 20 have also been confirmed by previous studies (Zhang, 2014), and the sites associated with chromosome 6, 9 are two new research intervals, which are of great significance for future studies, The related genes of this site should be explored and studied.

Furthermore, sequence comparison of genes within the remaining three significant correlation sites allowed us to predict 12 candidate genes closely linked to the regulation of salt tolerance during the germination stage of soybeans. These candidate genes play roles in the coordination of cellular responses, the regulation of osmotic stress, the attenuation of oxidative stress, the clearance of reactive oxygen species (ROS), and the management of heavy metal ion transport. Collectively, these genes are vital components in plant development, stress tolerance, and the maintenance of normal growth, immune response, and tolerance to abiotic and biotic stresses.

TABLE 5 Specific primers for KASP.

Primer name		Primer sequences
S05_41921861	F1	**GAAGGTGACCAAGTTCATGCT**GTATAAAGTTGAGGACTG**C**
	F2	**GAAGGTCGGAGTCAACGGATT**GTATAAAGTTGAGGACTG**A**
	R	TGGTGCTGACTTAGGCACTG
S02_6088007	F1	**GAAGGTGACCAAGTTCATGCT**TATTAATTTATTATTTTTT**G**
	F2	**GAAGGTCGGAGTCAACGGATT**TATTAATTTATTATTTTTT**A**
	R	TAGCAATGGCATGCACCTCA

The bold text stands for Fluorescent junction sequence.

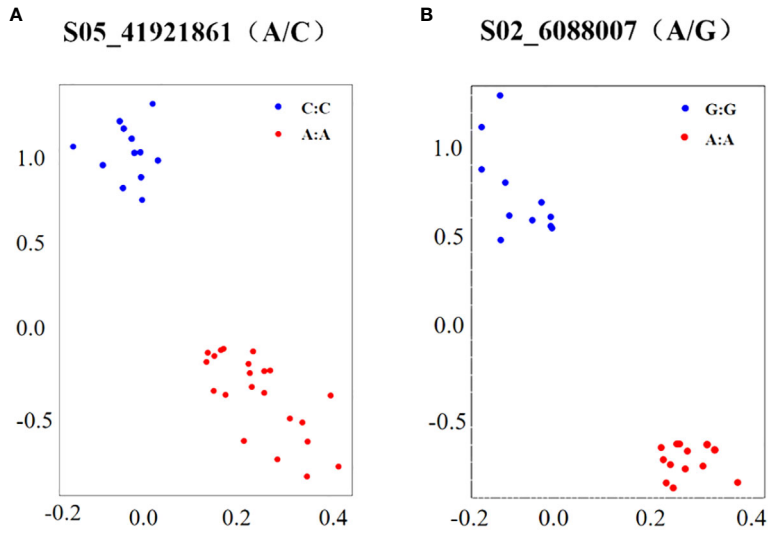


FIGURE 5
Genotyping of KASP markers. **(A, B)** are genotyping S05_41921861 and S02_6088007, respectively; X-axis and Y-axis scales are the values of the transmitted fluorescence, respectively.

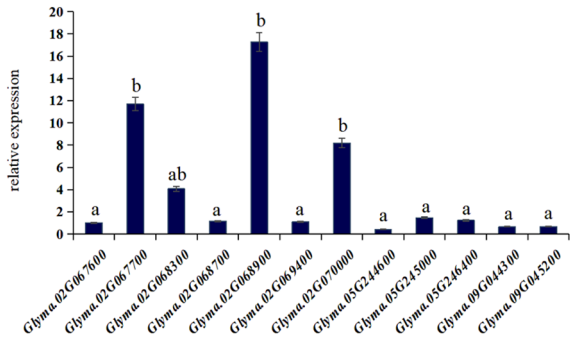


FIGURE 6
The expression levels of candidate genes. a, b represent significant differences.

Our findings contribute valuable genetic resources and a solid theoretical foundation for the breeding of salt-tolerant soybeans. They represent a critical step towards addressing the challenges of saline-alkali soybean production and increasing soybean yield, thereby bolstering food security.

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

WJ: Writing – original draft, Writing – review & editing. MZ: Writing – review & editing. HZ: Writing – review & editing. XL: Writing – review & editing. WZ: Writing – review & editing. QW: Writing – review & editing. JQ: Writing – review & editing. DX: Writing – review & editing. HC: Writing – review & editing. CS: Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. the National Natural Science Foundation of China (32001455), the Jiangsu Agriculture Science and Technology Innovation Fund (CX(23)1019), the Natural Science Foundation of Shandong Province of China (ZR2021MC071), the National Key Research and Development Program (2022YFD2300101-1), the Seed-Industrialized Development Program in Shandong Province (2021LZGC003), Qingdao Science and Technology Benefit the People Demonstration Project (23-2-8-xdny-10-nsh),the International Cooperation Project of Jiangsu Academy of Agricultural Sciences,Zhong shan Biological Breeding Laboratory (ZSBBL).

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

References

Bian, S. D., Sun, G., Shan, B., Zhou, H., Wang, J., Li, X., et al. (2020). Characterization of the soybean R2R3-MYB transcription factor GmMYB81 and its functional roles under abiotic stresses. *Gene* 753, 144803. doi: 10.1016/j.gene.2020.144803

Cheon, K. S., Jeong, Y. M., Oh, H., Oh, J., Kang, D. Y., Kim, N., et al. (2020). Development of 454 new kompetitive allele-specific PCR (KASP) markers for temperate japonica rice varieties. *Plants Basel* 9 (11), 1531. doi: 10.3390/plants9111531

De Silva, J., Jarman, C. D., Arrowsmith, D. A., Stronach, M. S., Chengappa, S., Sidebottom, C., et al. (1993). Molecular characterization of a xyloglucan-specific endo-(1→4)-β-d-glucanase (xyloglucan endo-transglycosylase) from nasturtium seeds. *Plant J.* 3 (5), 701–711. doi: 10.1046/j.1365-313X.1993.03050701.x

Ertiro, B. T., Ogugo, V., Worku, M., Das, B., Olsen, M., Labuschagne, M., et al. (2015). Comparison of kompetitive allele specific PCR (KASP) and genotyping by sequencing (GBS) for quality control analysis in maize. *BMC Genomics* 16 (1), 1–12.

Gao, C. S., Yuan, J., Zhi, J., Cheng Dong, Y. L., and Cheng, Q. (2023). Identification and acclimation analysis of ERF salt tolerance genes in soybean. *Journal of Plant Genetic Resources* (01), 30-38. doi: 10.13430/j.cnki.jpgr.20230510002

Hamwieh, A., Tuyen, D. D., Cong, H., Benitez, E. R., Takahashi, R., and Xu, D. H. (2011). Identification and validation of a major QTL for salt tolerance in soybean. *Euphytica* 179 (3), 451–459. doi: 10.1007/s10681-011-0347-8

Hao, X. F., Gao, H. X., Yan, P. M., et al. (2013). Salt stress effect on seed germination and physiological of soybean. *J. hubei Agric. Sci.* 52 (6), 1263–1266. doi: 10.14088/j.carolcarrollnkiissn0439-8114.2013.06.050

He, Z., Cui, Y. c., Xu, Q. g., Mao, D. h., et al. (2023). Genome-wide association analysis of anthocyanin content in rice seed pericarp. *Agric. modernization Res.* 44.02, 370–380. doi: 10.13872/j.1000-0275.2023.0027

Huang, G. T., Ma, S. L., Bai, L. P., et al. (2012). Signal transduction during cold, salt, and drought stresses in plants. *Mol. Biol. Rep.* 39 (2), 969–987. doi: 10.1007/s11033-011-0823-1

Jiang, P., Zhang, P., Wu, L., He, Y., Li, C., Ma, H., et al. (2021). Linkage and association mapping and kompetitive allele-specific PCR marker development for improving grain protein content in wheat. *Theor. Appl. Genet.* 134 (11), 3563–3575. doi: 10.1007/s00122-021-03913-z

Kan, G., Ning, L., Li, Y., Hu, Z., Zhang, W., He, X., et al. (2016). Identification of novel loci for salt stress at the seed germination stage insoybean. *Breed Sci.* 4, 530–541. doi: 10.1270/jsbbs.15147

Kan, G., Zhang, W., Yang, W., Ma, D., Zhang, D., Hao, D., et al. (2015). Association mapping of soybean seed germination under salt stress. *Mol. Genet. Genomics* 290 (6), 2147–2162. doi: 10.1007/s00438-015-1066-y

Kun Dziayi Turhan, Y. A., Midjiti, A. Y., Yongliang, Y., Yi, R., Xiaolei, S., and Hongwei, G. (2021). Genome-wide association analysis of salt tolerance in wheat during germination. *J. Xinjiang Agric. Univ.* 44 (02), 79–90. doi: 10.27431

Li, F. (2011). Soybean processing and utilization in our country development research. *J. Agric. Sci. Technol. Equip.* 01), 6–8. doi: 10.16313/j.cnki.nykjyzb.2011.01.004

Liang, T. Y., Gu, Y. Z., Ma, Y. J., Wang, H., Yang, G., and Qiu, L. J. (2023). Soybean resistance to low phosphorus genome-wide association analysis. *J. Plant Genet. Resour.* 24 (01), 237–251. doi: 10.13430/j.carolcarrollnkiJPGR.20220721003

Lipka, A. E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P. J., et al. (2012). GAPIT: genome association and prediction integrated tool. *Bioinformatics* 28 (18), 2397–2399. doi: 10.1093/bioinformatics/bts444

Liu, D. W., Ghosh, D., and Lin, X. H. (2008). Estimation and testing for the effect of a genetic pathway on a disease outcome using logistic kernel machine regression via logistic mixed models. *BMC Bioinf.* 9, 292. doi: 10.1186/1471-2105-9-292

Lu, Z. H., Ning, Z. L., Shi, B. W., Fan, Y. F., Yang, W. Y., and Yang, F. (2023). Response of leaf morphology and photosynthetic characteristics to shading in wild and cultivated soybeans. *J. Oil Crops* 45 (6), 1295–1304. doi: 10.19802/j.iSSN.1007-9084.2022338

Luan, J. H., Song, X. Y., Wang, L., Sun, L. L., Cheng, Y. S., Dong, H., et al. (2023). Differences Research in Salt Tolerance of New Rice Lines at Seedling Stage iniaoning. *Crop J.* 39 (3), 20–26.

Ma, J., Liu, J. b., and Zhu, W. h. (2023). Genome-wide association study and prediction for combining ability of maize agronomic traits. *J. Nucl. Agric.* 37 (05), 944–956.

Mao, Q. L., Wang, S., and Hu, B. (2020). Establishment of salt-alkali tolerant identification system and selection of salt-alkali tolerant germplasm in rapeseeds (Brassica napus L.). *Chin. J. Oil Crop Sci* 45 (4), 776. doi: 10.14088/j.cnki.issn0439-8114.2020.S1.085

Meng, N., Wei, M., and Wei, S. H. (2021). Effects of chlorine channel inhibitor Zn~(2+) on leaf traits of cultivated soybean under salt stress. *J. Plant Physiol.* 57 (10), 1955–1962. doi: 10.13592/j.carolcarrollnkiPPJ.2021.0164

Teng, W. L., Fu, X., Liu, J., Zhang, X. ch., Shi, F. f., Wang, Bo., et al. (2022). Mapping of additive and epistatic QTLs for salt tolerance traits of soybean RIL population at germination stage. *J. Northeast Agricultural University* (04), 1–8. doi: 10.19720/j.cnki.issn.1005-9369.2022.04.001

Tian, Y., Yang, L., Li, Y. H., and Qiu, L. J. (2018). Development and utilization of KASP markers at SCN3-11 sites of soybean cystocystitis elegans. *Acta Agronomica Sin.* 44 (11), 1600–1611. doi: 10.3724/SP.J.1006.2018.01600

Wang, F. B., and Jia, C. P. (2019). Identification of drought resistance ofmain rice varieties at bud stage in Xinjiang. *Mol. PlantBreeding* 17 (16), 5398–5405.

Wang, D. G., Yang, Y., Hu, G. Y., and Huang, Z. P. (2023). Comparative analysis of yield traits of soybean lines (species) in southern Huang-Huai region from 2016 to 2021. *Chin. J. oil-bearing Crops*, 1–9. doi: 10.19802/j.iSSN.1007-9084.2022256

Xiao, X. J., Chen, M., Han, D. P., Yu, R. L., Zheng, W., Xiao, G. B., et al. (2023). Bnapus rape seed number per pod genome-wide association analysis. *J. Biotechnol. Bull.* 33 (3), 143–151. doi: 10.13560/j.carolcarrollnkibiotech.Bull.1985.20220824

Xu, F. F., Chu, J. Y., Liu, Y., Xu, P. h., and Zhi, T. (2017). Effects of salt stress on water absorption and hydrolase activities of soybean seeds during germination. *Soybean Sci.* 36 (01), 74–77.

Yang, H., Xiang, S. H., Liu, L., Yang, X., Shu, Y. J., and He, Q. Y. (2023). Each soybean growth period traits of genome-wide association analysis. *J. Crops* 49 (10), 2727–2742.

Yu, D. Y., Huang, F., Ka, G. Z., Du, H. Y., Wang, H., et al. (2019). A cation diffusion facilitator, GmCDF1, negatively regulates salt tolerance in soybean. doi: 10.1371/journal.pgen.1007798

Yuan, J. Z., Gao, C. S., Zhi, J. Y., Cheng, Y. H., Cheng, Q., and Dong, L. D. (2019). Identification of NAC transcription factors in response to salt stress in soybean. *Soybean Sci.* 42 (06), 664–673.

Yuan, L., Wang, H., Wang, Q. L., Zhu, S. D., Zhao, Y., Yuan, M. M., et al. (2023). Shade maize inbred lines under the condition of combining ability and genetic effect analysis. *Crops* 39 (4), 104–109. doi: 10.16035/j.iSSN.1001-7283.2023.04.016

Zhai, H., Wan D, L., Yang, J. Y., Zhang, F., Wang, R. G., and Wan, Y. Q. (2023). Research progress on salt tolerance of lemon sticks Molecular Plant Breeding 1–12. Available at: http://kns.cnki.net/kcms/detail/46.1068.S.20230427.1336.006.html.

Zhang, M., Cao, Y., Wang, Z., Wang, Z. Q., Shi, J., Liang, X. Y., et al. (2018). Retrotransposon in an HKT1familsodium transporter causes variation of leaf Natexclusion and salt tolerance in maize. *New Phytol.* 217 (3), 1161–1176. doi: 10.1111/nph.14882

Zhang, W. J., Niu, Y., Bu, S. H., Li, M., Feng, J. Y., Zhang, J., et al. (2014). Epistatic association mapping for alkaline and salinity tolerance traits in the soybean germination stage. *PloS One* 9 (1), e84750. doi: 10.1371/journal.pone.0084750

Zhang, Y. H. (2014). *Genetic dissection of seed traits of chinese soybean landrace population and Its utilization in breeding by design Nanjing Agricultural University 2014.*

Zhang, W., Xu, W., Zhang, H., Liu, X., Cui, X., Li, S., et al. (2021). Comparative selective signature analysis and highresolution GWAS reveal a new candidate gene controlling seed weight in soybean. *Theor. Appl. Genet.* 134 (5), 1329 1341. doi: 10.1007/s00122-021-03774-6

Zhao, Q. Y., Xia, Y., and Zou, B. D. (2022). Agriculture and technology. *Soil salinization cause harm and recovery* 42, 11, 115–119. doi: 10.19754/j.nyyjs.20220615030

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 15 February 2024
DOI 10.3389/fpls.2024.1352379

OPEN ACCESS

EDITED BY
Peng Wang,
Jiangsu Province and Chinese Academy of Sciences, China

REVIEWED BY
Yingpeng Han,
Northeast Agricultural University, China
Yongguang Li,
Northeast Agricultural University, China

*CORRESPONDENCE
Xin Sun
18796799030@163.com
Huatao Chen
cht@jaas.ac.cn

RECEIVED 08 December 2023
ACCEPTED 30 January 2024
PUBLISHED 15 February 2024

CITATION
Jia Q, Zhou M, Xiong Y, Wang J, Xu D, Zhang H, Liu X, Zhang W, Wang Q, Sun X and Chen H (2024) Development of KASP markers assisted with soybean drought tolerance in the germination stage based on GWAS. *Front. Plant Sci.* 15:1352379. doi: 10.3389/fpls.2024.1352379

Development of KASP markers assisted with soybean drought tolerance in the germination stage based on GWAS

Qianru Jia[1], Miaomiao Zhou[1], Yawen Xiong[1,2], Junyan Wang[1], Donghe Xu[3], Hongmei Zhang[1], Xiaoqing Liu[1], Wei Zhang[1], Qiong Wang[1], Xin Sun[1*] and Huatao Chen[1,4*]

[1]Institute of Industrial Crops, Jiangsu Academy of Agricultural Sciences, Nanjing, China, [2]College of Life Science, Nanjing Agricultural University, Nanjing, China, [3]Japan International Research Center for Agricultural Sciences (JIRCAS), Tsukuba, Japan, [4]Zhongshan Biological Breeding Laboratory (ZSBBL), Nanjing, China

Soybean [*Glycine max*(L.)Merr.] is a leading oil-bearing crop and cultivated globally over a vast scale. The agricultural landscape in China faces a formidable challenge with drought significantly impacting soybean production. In this study, we treated a natural population of 264 Chinese soybean accessions using 15% PEG-6000 and used GR, GE, GI, RGR, RGE, RGI and ASFV as evaluation index. Using the ASFV, we screened 17 strong drought-tolerant soybean germplasm in the germination stage. Leveraging 2,597,425 high-density SNP markers, we conducted Genome-Wide Association Studies (GWAS) and identified 92 SNPs and 9 candidate genes significantly associated with drought tolerance. Furthermore, we developed two KASP markers for S14_5147797 and S18_53902767, which closely linked to drought tolerance. This research not only enriches the pool of soybean germplasm resources but also establishes a robust foundation for the molecular breeding of drought tolerance soybean varieties.

KEYWORDS

soybean, drought tolerance, germination stage, GWAS, KASP

Introduction

Soybean, which originated in China, has a cultivation history spanning over 3000 years (Sedivy et al., 2017; Li, 2021) and has evolved into a globally embraced crop due to its valuable composition. Soybean functions not only as a primary source of plant protein and oil for human consumption but also as fodder for animals (Zhao et al., 2021). Soybean also stands out for its unique contribution as a natural nitrogen fertilizer, thanks to its distinctive nodule structure (Sun et al., 2023).

Drought stress has always constrained the agriculture development and affects plant multiple physiological and biochemical indexes at all growth stages such as seed germination, seedling development, and flowering time (Zia et al., 2021; Poudel et al., 2023). Excessive

drought disrupts photosynthesis, hampers metabolism, and ultimately jeopardizes crop survival (Ouyang et al., 2022). As global climate changes and human development, droughts will be more severe, frequent, and longer lasting in the future. Soybean is more sensitive to drought than other crops especially in germination stage (Zhao et al., 2022). Under dry or drought conditions, soybean yield may plummet by over 50%, highlighting the imperative to enhance soybean drought tolerance for food security (Arya et al., 2021; Cotrim et al., 2021). Given that seed germination marks the onset of plant growth and development (Waterworth et al., 2015), it is crucial to identify drought-tolerant genotypes during this stage for effective plant breeding.

There are numerous factors influence soybean seed germination, encompassing both external environmental conditions and internal seed factors. Among them, moisture is an important factor affecting germination. Exposure soybean to drought stress during germination stage can result in a 20% reduction in seedling numbers and a staggering 50% decrease in yield (Devi et al., 2014; Zhao et al., 2017). This poses a huge threat to agricultural production. Encouragingly, technological advancements now enable the screening of drought-tolerant varieties through methods such as molecular marker-assisted breeding and genetic modification.

With the rapid development of high throughput sequencing technology, genome-wide association studies (GWAS) have been widely used for plant genetic analysis. Notably, numerous single nucleotide polymorphisms (SNPs) and quantitative trait loci (QTLs) linked to drought tolerance in soybean have been identified through this approach (Hacisalihoglu et al., 2018). Recent investigations by Zhang et al. uncovered a significant drought-related locus on chromosome 16 (32,206,964 bp to 32,458,483 bp), and obtained a gene *Glyma.16G164400* (*GmPrx16*) associated with drought tolerance through haplotype analysis (Zhang et al., 2023). Saleem et al. conducted short- and long- duration drought experiments on a 359 soybean accessions at the seedling stage, and identifying 17 and 22 significant SNPs, respectively (Saleem et al., 2022). However, previous investigations into soybean drought tolerance primarily concentrated on seedling or mature stages, leaving uncertainty about whether the identified QTLs and SNPs exhibit similar adaptations during the germination stage. In contrast, Zhao et al. identified 26 SNPs on chromosomes 1, 4, 5, 6, 8, 9, 11, 15, and 20 associated with drought tolerance during the germination stage. Notably, this research utilized a multi-faceted evaluation approach, employing RGR (relative germination rate), RGE (relative germination energy), GDTI (germination drought tolerant index), GSI (germination stress index), and MFV (membership function value) as indicators (Zhao et al., 2022). However, due to differences of genetic background and environmental conditions, stable QTLs have not been detected under different environmental conditions, and valuable QTLs that could be further validated are also rare.

Building upon these advancements, our present study utilized Polyethylene glycol-6000 (PEG-6000) to modulate drought tolerance and determined the optimal treatment concentration through a concentration gradient. We assessed various germination traits, including germination rate (GR), germination energy (GE), germination index (GI), relative germination rate (RGR), relative germination energy (RGE), relative germination index (RGI), and average subordinative function value (ASFV) within a natural population of 264 Chinese soybean accessions during the germination stage. In addition, we conducted GWAS based on 2,597,425 high-density SNP markers, and identified SNPs significantly associated with drought tolerance. Subsequently, two KASP markers were successfully developed as a result of this comprehensive analysis. The overarching objective of this research is to screen and identify soybean germplasm resources that exhibit resilience to drought conditions.

Materials and methods

Plant materials and growth conditions

A natural population of 264 Chinese soybean accessions (with 212 improved varieties and 52 landraces) (Zhang et al., 2021) were used as the materials. All of the soybeans were planted in 2021 (E1) and 2022 (E2) in Nanjing, Jiangsu Province. Seeds in E1 and E2 were collected to conduct subsequent drought treatment experiments.

Drought treatment and phenotypic determination

Dry soybean seeds with full grains, complete seed coat and uniform size were selected and disinfected with 2% NaClO2 for 15 min and then rinse three times with sterile water. Fifty of them put in a square with the size of 200 × 70× 70 mm. In each square, 30 g vermiculite is laid under the soybean seeds as a germination bed, and the surface of the seeds is covered with double layer filter paper to keep the surface of the seeds moist. A series of concentration gradients 0, 5%, 10%, 15%, 20% and 25% (w/v) of PEG-6000 (polyethylene glycol-6000) were used to simulate drought stress. After determining the optimal concentration, 60 mL of 15% (w/v) PEG-6000 was added to each soybean accession, and pure water treatment was used as the control. The temperature was set at 25 ± 1 °C, and the experiment was conducted under dark condition. Each processing is repeated three times. The germination was defined as when the length of the embryonic root extending beyond the hilum exceeding half of the longitudinal length of the seed,. The number of sprouts was counted daily. After the embryonic root breaks through the seed coat, 7 days is considered effective germination. GR, GE, GI, RGR, RGE and RGI were measured through the following formulas:

$$GR(\%)=(\text{Number of germinated seeds on the seventh day}/\text{total number of seeds})\times 100;$$

$$GE(\%)=(\text{Number of germinated seeds on the fourth day}/\text{total number of seeds})\times 100;$$

$$GI=\Sigma(Gt/Dt);$$

$$RGR=R_D/R_C;$$

$$RGE=E_D/E_C;$$

$$RGI=I_D/I_C.$$

In the above formulas, Gt is the number of germinated seeds per day, Dt is the number of germination days, D is the drought stress group, and C is the control group (Li et al., 2022).

Heritability (h^2) calculation following the formula:

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

σ_g^2 (g=1,2,3… 264) is the genotype variance of the test material, σ_{ge}^2 (e=1,2) is the variance of the interaction between the genotype and the environment of the test material, σ^2 is the error variance, n is the number of environments, and r is the number of replicates (Knapp et al., 1985).

To calculate the ASFV, we used the following series of formulas calculations (Liu et al., 2005).

(1) Firstly, we calculated the drought tolerance coefficient (DTC) of each genotype;

$$DTC_{ij} = \bar{y}_{ij(treatment)}/\bar{y}_{ij(Control)} \times 100$$

$\bar{y}_{ij(treatment)}$ and $\bar{y}_{ij(Control)}$ represent the j (j=1, 2, 3) trait average observed values of genotype i (i=1, 2, 3… 264) under drought and control treatment, respectively.

(2) We then standardized the DTC of each genotype using the subordinate function value (SFV);

$$F_{ij} = (DTC_{ij} - \min(DTCij))/(\max(DTC_{ij}) - \min(DTC_{ij}))$$

min(DTC_{ij}) and max(DTC_{ij})represent the j (j=1,2,3) trait minimum and maximum DTC of genotype i (i=1,2,3… 264), respectively;

(3) Lastly, we calculated the average drought resistance coefficient of three drought- related traits, denoted as the ASFV.

$$A(F_{ij}) = \frac{1}{3}\sum_{j=1}^{3} F_{ij}$$

Each genotype was assigned to a specific category corresponding to its level of drought tolerance, with the principle that a higher ASFV indicates a stronger drought tolerance of the genotype.

GWAS

The SNP markers used for whole GWAS in soybean natural populations were derived from pre laboratory resequencing work, resulting in a high-density physical map encompassing a total of 2,597,425 SNPs (Zhang et al., 2021). GWAS was performed using the GAPIT package based on R software (Wang and Zhang, 2021). To mitigate the risk of false-positive associations, a mixed linear model (MLM) was implemented for GWAS (Wang et al., 2010). The significance threshold for identifying association sites was set at $-Log_{10}(P) \geq 5.0$. Any SNP surpassing this threshold was deemed a significant association site.

KASP

Genotyping was carried out using three sets of primers (F1, F2, and R) specifically designed for the KASP markers of S14_5147797 and S18_53902767 (Supplementary Table S3). These primers were designed using the Primer-Blast tool available on the NCBI website (https://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi?LINK_LOC=BlastHome). Genomic DNA were extracted using the method of 2×CTAB (Mavrodiev et al., 2021). The PCR was amplified using the KASP V4.0 2×Mastermix (LGC, England), and the amplified procedure was conducted following the reagent instruction using a Quantitative Real-Time PCR System (ABI7500).

Quantification and statistical analysis

P value, F value and degree of freedom were calculated by the IBM SPSS Statistics 25 software (IBM, Armonk, NY, USA). Means were compared using a one-way analysis of variance (ANOVA). The heritability was calculated using the lme4 package in R (http://www.Rproject.org/). A frequency distribution was created using Microsoft Excel 2021. To ensure statistical significance, analyses were performed in sufficiently large samples in all experiments.

Results

Screening of optimal concentration of PEG-6000

Six soybean germplasm (NPS36, NPS137, NPS140, NPS196, NPS213, NPS233) were selected randomly from the 264 Chinese soybean accessions. Germplasms underwent individual treatment with varying concentrations of PEG-6000, namely 0%, 5%, 10%, 15%, 20%, and 25% (w/v) PEG-6000 respectively. Six days later, the GR, GE and GI were assessed to identify the optimal concentration of PEG-6000. The results showed a consistent decline in GR, GE and GI with increasing concentrations of PEG-6000 for each soybean germplasms (Supplementary Figures S1A, B). Under 5% and 10% PEG-6000 treatment, GR, GE and GI exhibited no significant difference from the control (0%). However, under 20% and 25% PEG-6000 treatment, GR, GE and GI of some germplasms went to zero (Supplementary Figures S1A–C), rendering these concentrations unsuitable for further research. Only at the 15% PEG-6000 treatment did all GR, GE, and GI values exhibited a statistically significant difference compared to the control. Consequently, 15% PEG-6000 was identified as the optimal concentration for subsequent experiments.

Descriptive statistics on the drought-tolerance traits of the 264 soybean accessions

This study investigated the drought-tolerance traits of soybean harvested in 2021 (E1) and 2022 (E2) during the germination stage.

We treated the 264 soybean accessions with 0% (control group) and 15% PEG-6000 (drought treatment) and measured the GR, GE, and GI in two environments (Table 1). GRs of the 264 soybean accessions under drought treatment in E1 and E2 were 50.13% and 51.10%, while the controls were 88.55% and 83.90%, respectively. The coefficient of variations (CVs) of the GR under drought treatment were 49.20% and 41.69% while the controls were 11.56% and 13.49%, respectively. The GEs under treatment in E1 and E2 were 25.60% and 19.98%, while in the control conditions, they were 81.31% and 71.28%. The CVs of GE under drought treatment were 78.88% and 76.71%, while in the controls they were 19.26% and 21.44% in E1 and E2, respectively; the GIs under treatment were 9.87% and 6.33%, while the controls they were 2.82% and 2.51%, respectively. In E1 and E2, the CVs of GI after treatment were 61.18% and 54.27%, while the controls were 29.04% and 16.10%, respectively (Table 1). The three basic indicators of the control group were significantly greater than that under drought treatment, indicating that after 15% PEG-6000 drought treatment, the GR of soybean germplasms was greatly inhibited, leading to a decrease in GR, GE, and GI. At the same time, the CV of all three traits under drought treatment was significantly greater than that of control, indicating that GI, GR, and GE responses to drought vary significantly.

Moreover, a descriptive statistical analysis was conducted on RGR, RGE, RGI, and ASFV of 264 accessions. The average values of RGR, RGE, RGI, and ASFV in E1 and were 0.56, 0.30, 0.28, 0.40, 0.61, 0.28, 0.39 and 0.38, respectively (Table 2). The CVs of RGR were 47.00% and 42.11%, respectively. The CVs of RGR and RGI in E1 and E2 were 74.90%, 72.39%, 56.18% and 51.48%, respectively. The CVs of ASFV were 52.86% and 49.25%, surpassing the 50% threshold in both cases. These indicate that the population used in this study exhibits rich variation under drought conditions. Furthermore, the statistical analysis revealed heritability values of RGR, RGE, RGI, and ASFV were 49.82%, 44.31%, 44.14%, and 49.11%, respectively (Table 2). To a certain extent, it is beneficial for the screening of drought resistant soybean germplasm in germination stage.

The variation amplitude of GR, GE, and GI of the 264 soybean accessions in E1 surpassed that of E2, with each trait exhibiting a higher CV in E1 compared to E2. Notably, GR, GE, and GI displayed significant differences between E1 and E2, indicating that the drought tolerance traits of soybean germination stage may be selected during the process of variety improvement. Among the three drought tolerance traits, the RGR, RGE and RGI variation range of E1 and E2 spanned from 0.00~1.04, 0.00~1.47, 0~1.00, 0.00~0.88, 0.00~0.78 and 0.00~0.97, respectively (Table 2). It is evident that the variation range of RGI is comparatively smaller than that of RGR and RGE in E1 and E2. These indicated that the drought-tolerant traits influenced by environment in the soybean germination stage. Additionally, there were notable variations among the 264 soybean accessions.

Variance and correlation analysis of drought tolerance traits across the 264 soybean accessions

We conducted a variance analysis on RGR, RGE and RGI of 264 soybean accessions in E1 and E2 (Table 3). The results showed that there were significant differences among genotypes, drought treatments and different environments. There are significant differences in genotype and environmental interaction effects between RGR and RGI, indicating that each genotype is influenced by environmental factors while experiencing differences in drought stress, and changes with environmental changes.

To explore the correlation among RGR, RGE and RGI in E1 and E2, correlation analysis was conducted (Figure 1). The results showed that, the correlation of RGR, RGE and RGI of E1 and E2

TABLE 1 Descriptive statistics of three germination-related traits of 264 accessions treated with 0% and 15% PEG-6000.

Trait	Treat		Max	Min	Range	Mean	SD	CV (%)
GR	E1	C	100.00.	41.67	41.67~100.00	88.55	10.24	11.56
		D	96.67	0.00	0.00~96.67	50.13	24.67	49.20
	E2	C	100.00	45.00	45.00~100.00	83.90	11.32	13.49
		D	90.00	0.00	0.00~90.00	51.10	21.30	41.69
GE	E1	C	100.00	33.33	33.33~100.00	81.31	15.66	19.26
		D	90.00	0.00	0.00~90.00	25.60	20.19	78.88
	E2	C	98.33	25.00	25.00~98.33	71.28	15.28	21.44
		D	75.00	0.00	0.00~75.00	19.98	15.33	76.71
GI	E1	C	3.861	17.111	3.86~17.11	9.87	2.87	29.04
		D	7.72	0.00	0.00~7.72	2.82	1.72	61.18
	E2	C	8.194	3.611	3.91~8.19	6.33	1.02	16 .10
		D	6.69	0.00	0.00~6.69	2.51	1.36	54.27

TABLE 2 Descriptive statistics of germination-related traits of 264 accessions under drought stress.

Trait	Environment	Max	Min	Mean		SD	CV (%)	Skewness	Kurtosis	H^2 (%)
RGR	E1	1.04	0.00	0.56	0.59	0.26	47.00	-0.231	-0.846	49.82
	E2	1.47	0.00	0.61		0.26	42.11	-0.026	-0.033	
RGE	E1	1.00	0.00	0.30	0.29	0.23	74.90	0.723	-0.146	44.31
	E2	0.88	0.00	0.28		0.20	72.39	0.684	-0.297	
RGI	E1	0.78	0.00	0.28	0.34	0.16	56.18	0.533	-0.031	44.14
	E2	0.97	0.00	0.39		0.20	51.48	0.265	-0.576	
ASFV	E1	0.94	0.00	0.40	0.39	0.21	52.86	0.206	-0.719	49.11
	E2	0.91	0.14	0.38		0.19	49.25	0.244	-0.579	

RGR, relative germination rate; RGE, relative germination energy; RGI, relative germination index; ASFV, average subordinative function value; SD, standard deviation; CV, coefficient of variation; H^2, heritability.

is poor (Figure 1A). However, positive correlations were observed among RGR, RGE, RGI, and ASFV, with correlation coefficients of 0.70, 0.86 and 0.91, respectively (Figure 1B). In particular, RGE exhibited a significant positive correlation with RGI and ASFV, with correlation coefficients of 0.86 and 0.92, respectively. Additionally, a significant positive correlation between RGI and ASFV was identified, with a correlation coefficient of 0.94 (Figure 1B). These results illuminated the intrinsic connections among various drought-tolerant traits.

Identification of drought tolerance grade and screening of drought-tolerance soybean germplasms

This study standardized the RGR, RGE, and RGI of E1 and E2 using the subordinative function method. The ASFV of each genotype in the natural population was calculated, which were categorized into five groups to establish distinct drought resistance grading standards (Figure 2). These categories were graded as high drought tolerance (HDT), drought tolerance (DT), medium drought tolerance (MDT), drought sensitive (DS) and high drought sensitive (HDS), where higher ASFV values indicated stronger drought resistance within soybean germplasms.

In E1, the ASFV or the categories were defined as follows: HDT (above 0.75), DT (0.56-0.75), MDT (0.38-0.56), DS (0.19-0.38), and HDS (below 0.19) (Figure 2A). Notably, MDT soybean germplasms were the most abundant, constituting approximately 32% of the total, while DS germplasms ranked second, comprising 27%. Drought-resistant and drought-sensitive soybean germplasms were relatively balanced. (Figure 2A). HDT germplasm accounts for 5.29% of the total natural population of soybeans.

In E2, the ASFV for the defined categories were as follows: HDT (above 0.73), DT (0.5-0.73), MDT (0.37-0.55), DS (0.18-0.37), and HDS (below 0.18) (Figure 2B). Among them, DS soybean germplasms were the most prevalent, constituting approximately 25%, and the MDT germplasms ranked second, representing 23%. The number of drought-resistant soybean germplasms was comparable to that of DS germplasms. HDS germplasm accounted for 2.27% of the total soybean accessions.

This study successfully identified a total of 17 drought-tolerant soybean germplasms (Table 4), characterized by high GR, GE and GI. In E1, the HDT soybean germplasms included NX-F7-13, NX-F4-4, NX-F4-3, Guichundou 113, Zhonghuang 306, Guichundou 111,

TABLE 3 Variance analysis of germination-related traits of 264 accessions under drought stress.

Trait	Variation sourse	DF	SS	MS	F value	P value
RGR	G	249	64.122	0.258	4.835	<0.05
	Env	1	0.479	0.479	8.995	<0.05
	G×Env	178	24.354	0.137	2.569	0.003
RGE	G	249	47.342	0.190	3.345	<0.05
	Env	1	0.073	0.073	1.287	<0.05
	G×Env	178	22.467	0.126	2.221	0.257
RGI	G	249	28.771	0.116	22.234	<0.05
	Env	1	1.317	1.317	253.360	<0.05
	G×Env	178	4.428	0.032	6.201	<0.05

G, germplasm; Env, environment; DF, degree of freedom; SS, sum of square; MS, mean square.

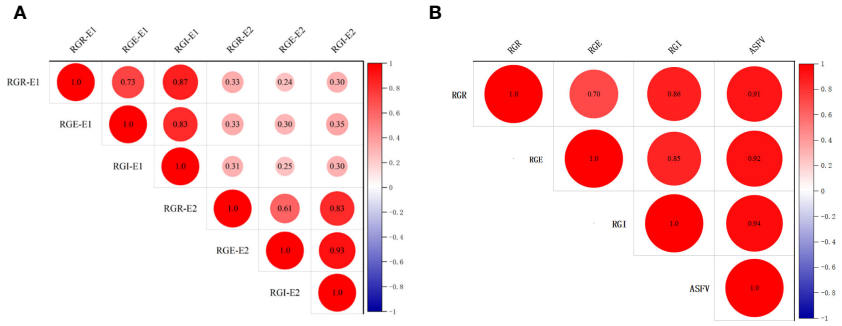


FIGURE 1
Correlation analysis of germination-related traits of 264 accessions. **(A)** Correlation analysis of RGR, RGE, RGI in E1 and E2. **(B)** Correlation analysis of RGR, RGE, RGI and ASFV.

Gui26BC2-7, Huaidou 5, Huaidou 1, Doujiao 73, Xu 0701, and Jining 98-11497 (Table 4). In E2, the HDT soybean germplasms comprised Zhou 11019-2-1, Guichun 16, Gui0508-3, Huaidou 1, Nannong15-3, and Qihuang 35 (Table 4). Notably, Huaidou 1 demonstrated high drought tolerance in both environments. These identified HDT materials within soybean accessions serve as a valuable research foundation for further exploration into the genetic mechanisms of soybean drought resistance and as experimental resources for future soybean drought-resistant breeding.

GWAS signals related to six germination traits

We calculated the frequency of RGR, RGE, and RGI of 264 soybean accessions in E1 and E2 using Microsoft Excel 2021, and drew frequency distribution maps and density curves (Supplementary Figure S2). The absolute values of RGR, RGE, and RGI kurtosis and skewness are less than 1 (Table 2) and the histograms of phenotypic data exhibit an approximate normal distribution (Supplementaryl Figure S2), which indicated that the natural soybean population in this study has rich genetic variation and is suitable for subsequent genome-wide association analysis.

To illuminate the genetic mechanism of six drought-tolerant germination traits, a GWAS with high-density SNPs was conducted to identify SNPs associated with six drought-tolerant germination traits. At a significance threshold of $-\text{Log}_{10}(P) \geq 5.0$, a total of 92 SNPs significantly associated with drought tolerance in germination stages were detected in this study, distributed across chromosomes 1, 2, 3, 4, 5, 6, 8, 9, 10, 11, 14, 16, 17, 18, 19 and 20 of the soybean genome (Figure 3, Supplementary Tables S1, S2).

Noteworthy SNPs included S08_37604504 locus on chromosome 8, significantly associated with GE, GI and RGE, explaining 10.82%, 10.08%, and 8.51% of the phenotypic variation, respectively (Supplementary Tables S1, S2). Additional SNPs like S18_32669137, S18_55108397 and S18_55108434 on chromosome 18 were significantly associated with the mentioned traits. S18_55108447 on chromosome 18 is significantly associated with GE, GI and RGE, explaining 10.2%, 10.07% and 10.2% of the phenotypic variation, respectively. On chromosome 9, S09_5096288 was significantly associated with GR and RGR, while S11_16409860 located on chromosome 11 was significantly associated with GE and GI. In addition, S11_2747816 and S11_2747823 on chromosome 11 were physically close to each other and related to RGE and RGI. S20_32847223 and S20_794406 located on chromosome 20, were significantly associated with GE and RGE. The SNPs significantly

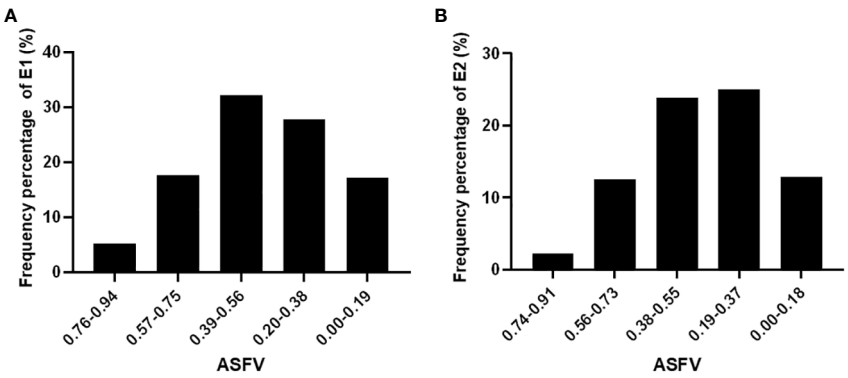


FIGURE 2
ASFV distribute frequency percentage in E1 **(A)** and E2 **(B)** environment.

TABLE 4 Drought-tolerant soybean germplasm screened in two environments.

Environment	High drought tolerance	ASFV
E1	NX-F7-13	0.79
	NX-F4-4	0.83
	NX-F4-3	0.79
	Guichundou 113	0.91
	Huaidou 5	0.82
	Huaidou 1	0.76
	Doujiao 73	0.80
	Xu 0701	0.80
	Jining 98-11497	0.94
	Zhonghuan 306	0.79
	Guichundou 111	0.79
	Gui 26BC2-7	0.86
E2	Huaidou 1	0.73
	Zhou 11019-2-1	0.74
	Guichun 16	0.78
	Gui 0508-3	0.78
	Nannong 15-3	0.75
	Qihuang 35	0.74

associated within 5543129~38193841 bp on chromosome 14 are all associated with two traits. The SNPs significantly associated with 35922040~36291124 bp on chromosome 17 are all related to GE.

Identification of candidate genes of soybean drought-tolerant traits

To investigate the phenotypic effects of allelic variations associated with significantly associated SNPs, haplotype analysis was conducted on SNPs surpassing the highest threshold. Notable findings include:

Allelic variation of S10_3999727 was A/C. Average RGR of germplasms carrying S10_3999727-A were higher compared to those carrying S10_3999727-C (Figure 4A).

Allelic variation of S14_5147797 was A/T, germplasms carrying S14_5147797-A exhibiting a significantly higher average RGE than those carrying S14_5147797-T (Figure 4B).

Allelic variation of S06_5126498 was C/T, germplasms carrying S06_5126498-C having a higher average RGI compared to those carrying S06_5126498-T (Figure 4C).

Allelic variation of S11_14316470 was T/C, germplasms carrying S11_14316470-C having lower RGR than those carrying S18_55033581-T (Figure 4D).

Allelic variation of S18_53902767 was A/G, average RGI of germplasms carrying S18_53902767-A were significantly lower than those carrying S18_53902767-T (Figure 4E).

Allelic variation of S18_55033581 was C/T, average RGI of germplasms carrying S18_55033581-C were significantly lower than that carrying S18_55033581-T (Figure 4F).

In a previous study, we had conducted a linkage disequilibrium (LD) analysis on the soybean natural population in which the LD increased to 120 kb (Zhang et al., 2021). Referring to the gene functional annotation information of soybean, nine candidate genes significantly associated with drought tolerance during soybean germination stage were identified (Table 5). Including *Glyma.06G065900* (serine-rich protein-related protein), *Glyma.06G066200* (TOPLESS-related 1), *Glyma.10G044200* (Acyl-CoA N-acyltransferase with RING/FYVE/PHD-type zinc finger domain), *Glyma.14G035500* (vesicle-associated membrane protein 713), *Glyma.14G035600* (CRT (chloroquine-resistance transporter)-like transporter 3), *Glyma.14G063700* (HSP20-like chaperones superfamily protein), *Glyma.18G264600* (membrane-anchored ubiquitin-fold protein 2) and *Glyma.18G266900* (glycosyl hydrolase family 81 protein). Additionally, according to transcriptome data, all these genes exhibited altered expression levels under drought conditions, either induced or reduced (Figure 5A) (Wang et al., 2021b). To verify candidate genes expression level in drought resistant and drought sensitive material, we used the NPS 62 as drought resistant material and NPS40 as drought sensitive material. After treated with 15% PEG-6000 8 h, we sampled and conducted a qPCR. As is shown, some genes of NPS62 response to drought stress levels were higher than that of NPS40 (*Glyma.10G044200*, *Glyma.14G035500 and Glyma.18G264600*). But some genes of NPS62 response to drought stress levels were lower than that of NPS40 (*Glyma.06G065900*, *Glyma.06G066200* and *Glyma.14G035600*). Also, some of them between two materials didn't show significant difference (*Glyma.14G063700, Glyma.18G252300 and Glyma.18G266900*) (Figure 5B).

Development of KASP markers of SNPs

Based on the GWAS results, we developed KASP markers for S14_5147797(A/T) and S18_53902767(A/G), which were significantly associated with RGE and RGI, respectively. Both of the genotypes can be clearly divided (Figure 6). In Figure 6A, blue dots represent soybean germplasm carrying the A allele mutation site, indicating higher RGE during the germination stage. Conversely, red dots represent soybean germplasm carrying the T allele mutation site, showing the opposite trend. In summary, soybean germplasm with the genotype AA generally exhibits a higher RGE than those with the genotype TT. In Figure 6B, blue dots represent soybean germplasms carrying the G allele mutation site, associated with a higher RGI during the germination stage. On the other hand, the red dot represents soybean germplasm carrying the A allele mutation site, correlated with a lower RGI. Overall, soybean germplasm with the genotype GG tends to have a higher RGI compared to those with the genotype AA. The distinct clustering of genotypes in these figures highlights the utility of the developed KASP markers for precise genotyping based on these significant SNPs, providing valuable information for further studies and breeding efforts targeting drought tolerance in soybeans during germination.

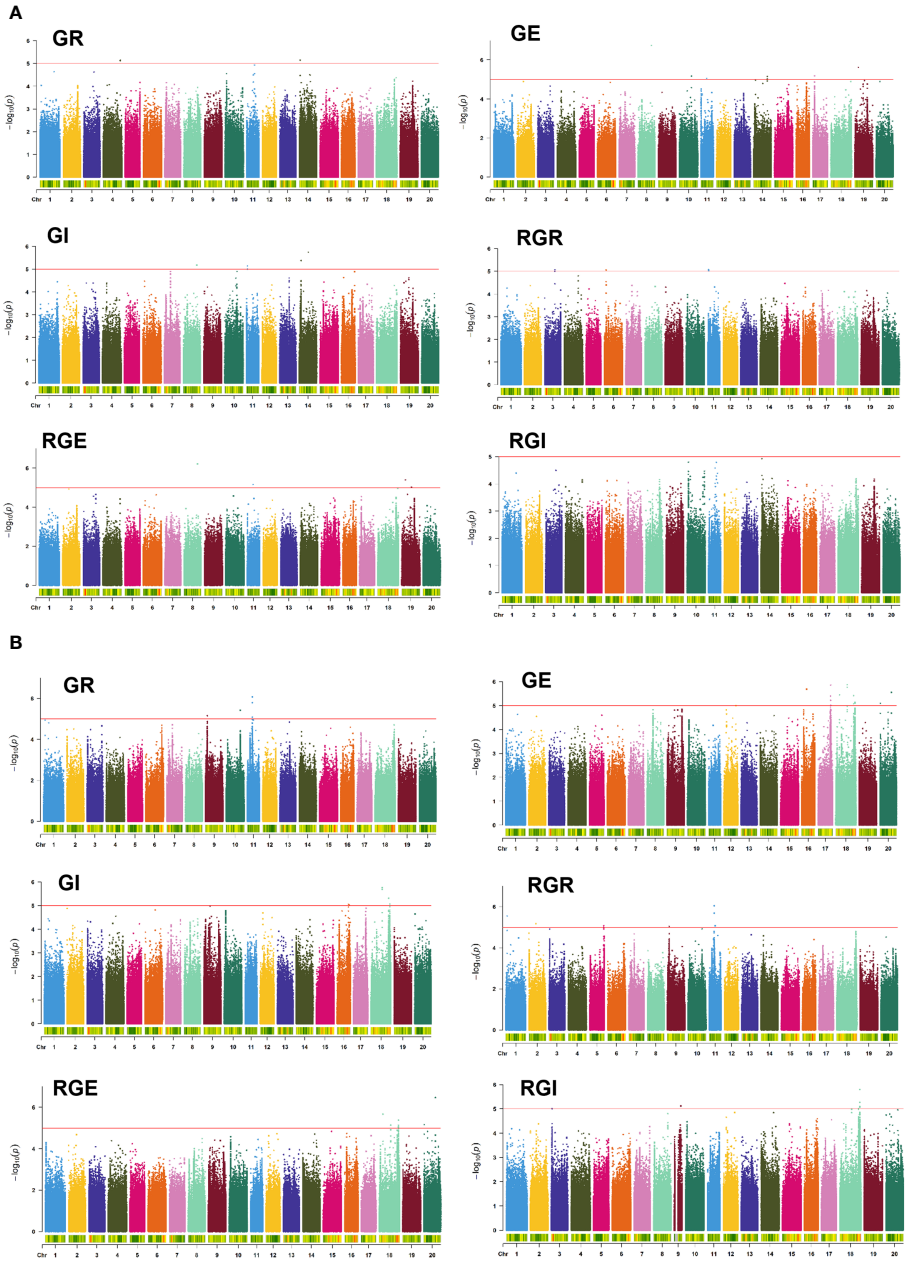


FIGURE 3
Manhattan plots for the GWAS for GR, GE, GI, RGR, RGE, and RGI in E1 **(A)** and E2 **(B)**. The red line indicates the significance threshold ($-\log_{10}(P)$=5.0).

Discussion

Drought tolerance remains a persistent challenge affecting crop growth, development, and yield. In the current study, we treated a natural soybean population of 264 accessions in the germination stage with15% PEG-6000 to modify the drought stress.

Soybean germplasms resistance to drought tolerance in the germination stage

Numerous studies have addressed the identification of drought resistance traits during soybean germination, employing diverse soybean populations and proposing varying indicators, leading to nuanced definitions (Yang et al., 2013). The choice of drought resistance indicators significantly influences the authenticity and reliability of experimental outcomes. Zhang et al. (2005) advocated for a comprehensive approach, combining physiological, biochemical and morphological indicators to assess and select drought-resistant varieties. Wang et al. (2012) argued that soybean varieties exhibiting high GE also have high GR, with a strong correlation between their membership function values and the total membership function values. Consequently, in evaluating drought resistance during soybean germination, GR and GE can serve as valuable indicators. Commonly used indicators for drought resistance during the germination period encompass RGR, RGE, RGI, and germination

10.3389/fpls.2024.1352379

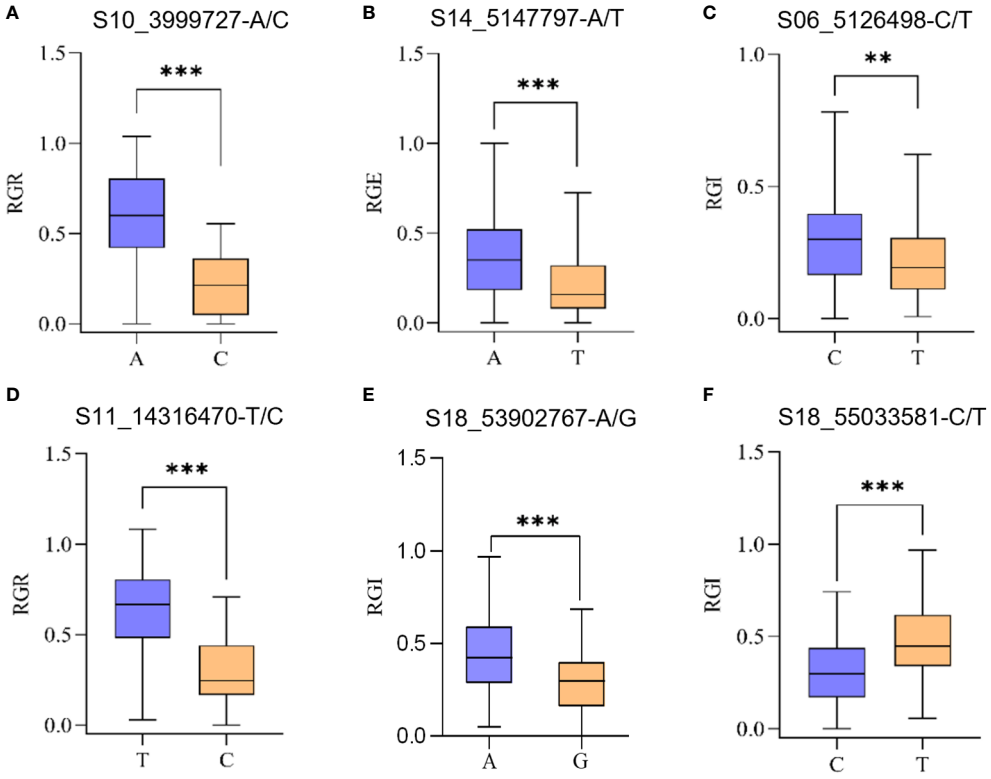


FIGURE 4
Significant SNP haplotype analysis of drought-related traits across 264 accessions. These SNPs were selected from the significant SNPs obtained from GWAS. **(A)** Average RGR of germplasms carrying S10_3999727-A/C; **(B)** Average RGE of germplasms carrying S14_5147797-A/T; **(C)** Average RGI of germplasms carrying S06_5126498-C/T; **(D)** Average RGR of germplasms carrying S11_14316470-T/C; **(E)** Average RGI of germplasms carrying S18_53902767-A/G; **(F)** Average RGI of germplasms carrying S18_55033581-C/T. The ** represents p<0.01 and *** represents p<0.001, one-way analysis of variance (ANOVA).

drought tolerance index (GDTI), among others (Li et al., 2010). Recently, Zhao et al. used RGR, RGE, GDTI, GSI and membership function value (MFV) as the drought tolerance indicators, evaluating a natural soybean population of 410 accessions with 158,327 SNPs. In alignment with these considerations, our study adopts GR, GE, and GI as primary evaluation indicators for soybean drought resistance during germination. Recognizing the complex nature of soybean drought resistance, this research employs RGR, RGE, and RGI as additional indicators, integrating a membership function method for a comprehensive assessment of drought resistance within this population during germination. This approach facilitates a thorough comparison across different genotypes, employing equal interval divisions based on the ASFV range to establish standardized grading criteria for drought tolerance. By reducing subjective human grouping, this methodology enhances the demonstration of soybean responses to drought stress. Our analysis of phenotypic data for drought tolerance traits in 264 natural soybean accessions during germination reveals significant genotype variations post-drought stress treatment, indicating a degree of genetic stability. This affirms the suitability of this population for genetic screening of drought-resistant soybean germplasm.

TABLE 5 Candidate genes for drought-tolerant in the germination stage.

Gene ID	Homologs	Functional annotation
Glyma.06G065900	*AT5G11090*	serine-rich protein-related
Glyma.06G066200	*AT1G80490*	TOPLESS-related 1
Glyma.10G044200	*AT2G37520*	Acyl-CoA N-acyltransferase with RING/FYVE/PHD-type zinc finger domain
Glyma.14G035500	*AT5G11150*	vesicle-associated membrane protein 713
Glyma.14G035600	*AT5G53540*	CRT (chloroquine-resistance transporter)-like transporter 3
Glyma.14G063700	*AT1G53540*	HSP20-like chaperones superfamily protein
Glyma.18G252300	*AT2G44840*	Ethylene-responsive element binding factor 13
Glyma.18G264600	*AT5G15460*	membrane-anchored ubiquitin-fold protein 2
Glyma.18G266900	*AT5G15870*	glycosyl hydrolase family 81 protein

This study categorized the natural soybean population into five distinct grades, ranging from HDT to HDS. A total of 17 soybean germplasms exhibiting HDT were selected (NX-F7-13, NX-F4-4, NX-F4-3, Guichundou 113, Zhonghuang 306, Guichundou 111, Gui26BC2-7, Huaidou 5, Huaidou 1, Doujiao 73, Xu 0701, Jining 98-11497, Zhou 11019-2-1, Guichun 16, Gui0508-3, Nannong 15-3

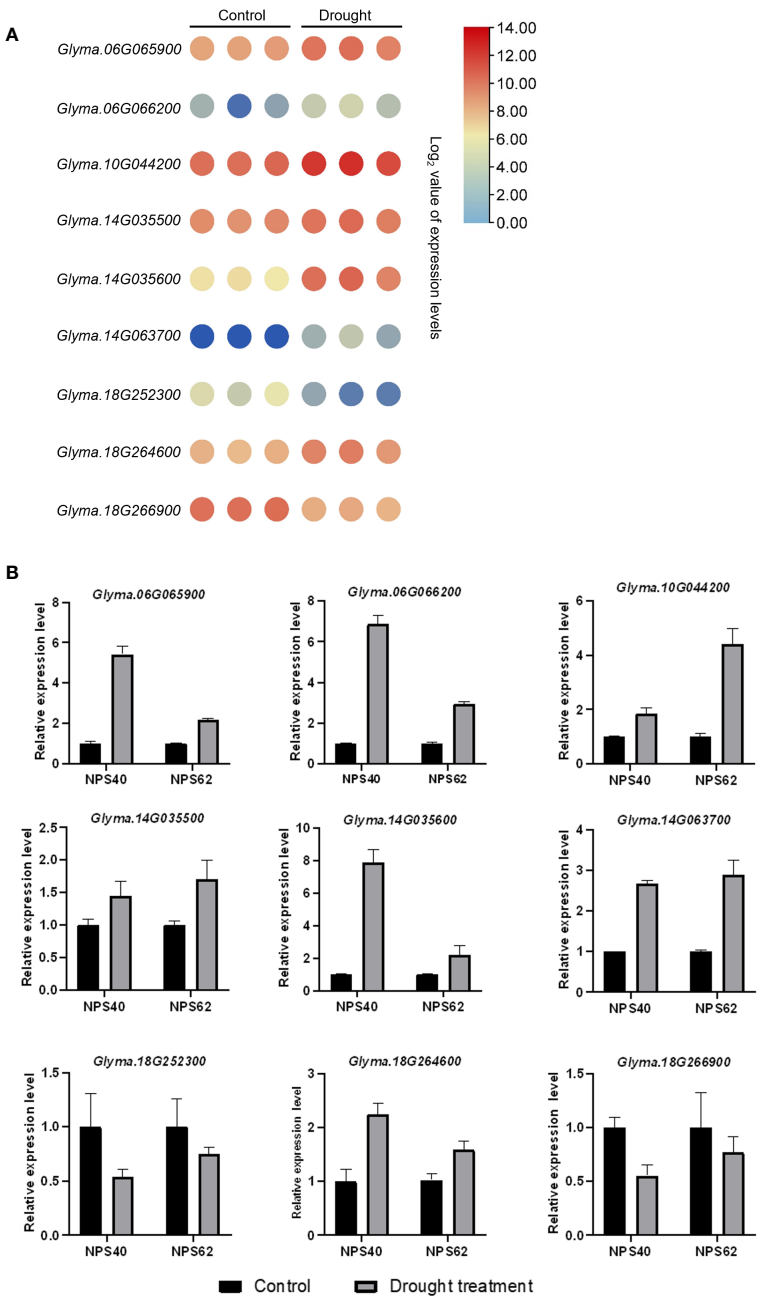


FIGURE 5
(A) Heat map of candidate genes relative expression levels under drought treatment. Each group has three replicates. One circle represents one value of the group. **(B)** Relative expression level of candidate genes of drought sensitive material NPS40 and drought tolerant material NPS 62.

and Qihuang 35). Remarkably, Huaidou 1 demonstrated HDT in both environmental conditions (Table 4). The selected germplasms serve as valuable resources for future in-depth investigations into the molecular mechanisms of soybean drought resistance and genetic breeding endeavors. It is crucial to note that this experiment artificially simulated drought condition, the data and conclusions obtained may not entirely substitute conclusions drawn from field soil environments. Furthermore, the experiment specifically focuses on the germination stage, and the identified conclusions may not extend to other stages such as seedling and maturity.

Candidate genes for resistance to drought tolerate in the germination stage

GWAS is widely used in the study of important agronomic traits. By effectively identifying SNP associated with the target trait, a total of 92 SNPs were associated with drought tolerance in the germination stage of this experiment (Figure 3, Supplementary Tables S1, S2). Unfortunately, the association SNPs identified in E1 and E2 did not co-locate with each other, perhaps indicating a lack of overlap between the two environments. This discrepancy may be attributed to variations in

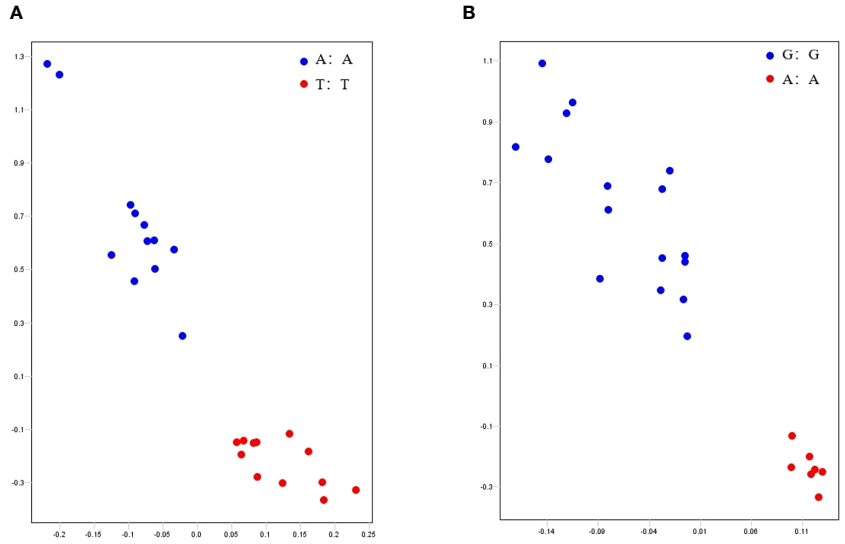


FIGURE 6
Genotyping of KASP markers. **(A, B)** were genotyping of S14_5147797 and S18_53902767, respectively. Blue and red dots in **(A)** represent the soybean germplasm carrying the A and T allele mutation sites, respectively. Blue and red dots in **(B)** represent the soybean germplasm carrying the G and A allele mutation sites, respectively.

environmental conditions and other factors influencing the genetic associations with drought tolerance during germination.

Based on the GWAS results, nine candidate genes were identified: *Glyma.06G065900*, *Glyma.06G066200*, *Glyma.10G044200*, *Glyma.14G035500*, *Glyma.14G035600*, *Glyma.14G063700*, *Glyma.18G252300*, *Glyma.18G264600*, *Glyma.18G266900*. *Glyma.06G065900* is a serine-rich protein-related protein. Serine metabolism plays a crucial role in the plant's response to various abiotic stresses (Kishor et al., 2020). When a plant is exposed to temperature, flooding, salt, drought or heat stress, serine accumulates (Stewart and Larher, 1980; Kaplan et al., 2004; Bocian et al., 2015; Hossain et al., 2017; Li et al., 2017). Transcriptome data also indicates an induction of *Glyma.06G065900* expression under drought stress (Figure 5). *Glyma.06G066200* codes a TOPLESS-related protein, TOPLESS (TPL) and TOPLESS-related protein (TPR) corepressors usually interact with transcription factors to regulate gene expression. TPRs have been known to influence hormonal signaling pathways, such as auxin, gibberellins, jasmonic acid and brassinosteroids, which are essential in plant stress responses (Saini and Nandi, 2022). Glyma.10G044200 is an Acyl-CoA N-acyltransferase with RING/FYVE/PHD-type zinc finger domain-containing protein. Acyl-CoA N-Acyltransferase involved in regulating plant meristem and architecture (Walla et al., 2020). Additionally, Acyl-CoA is the active metabolic intermediate in fatty acid synthesis and decomposition and potentially contributes to drought stress response through its involvement in lipid metabolism. *Glyma.14G035500* codes the protein vesicle-associated membrane protein 713. The exact role of this protein in drought response is not known, but vesicle-associated membrane proteins are generally involved in membrane trafficking, so it may be implicated in stress response pathways. From the web of TAIR (https://www.arabidopsis.org/index.jsp), we know that its homologous AtVAMP713 (AT1G53540) is involved in response to salt stress. A vesicle-associated membrane protein TaVAMP in wheat was identified as a drought-inducible protein, suggesting a potential role in abiotic stress tolerance (Singh et al., 2018). Glyma.14G035600 is a CRT-like transporter. In rice, OsCLT1 is a CRT-like transporter required for glutathione homeostasis and arsenic tolerance (Yang et al., 2016). The triple mutant *clt1 clt2 clt3* of Arabidopsis showed increased cadmium (Cd) sensitivity (Maughan et al., 2010). Further research is needed to determine whether it is involved in drought stress in soybean. Glyma.14G063700 and AT1G53540 (HSP17.6C) are homologous. It is involved in response to heat, hydrogen peroxide and salt stress (Yamaguchi et al., 2021; TAIR). *Glyma.18G252300* codes an ethylene-responsive transcription factor (ERF), which has been shown to be involved in responding to a wide range of abiotic stresses (Licausi et al., 2013; Xie et al., 2019). In wheat, TaERF87 and TaAKS1 synergistically regulate TaP5CS1/TaP5CR1-mediated proline biosynthesis to enhance drought tolerance (Du et al., 2023), while the transcriptome data reveals a reduction in the expression level of *Glyma.18G252300* in soybean under drought stress (Figure 5). Further experiments are required to elucidate the mechanisms underlying this observation. Glyma.18G264600 is a membrane-anchored ubiquitin-fold protein located in the plasma membrane. The plasma membrane is the primary site for sensing extracellular stimuli, and when cells are stimulated by abiotic stress, the cell membrane generates secondary signaling molecules such as reactive oxygen species (ROS) and phospholipids. Glyma.18G264600 may interact with second messengers to respond to drought stress. *Glyma.18G266900* is a glycosyl hydrolase family 81 protein, known to participate in responses to various stimuli such as bacterium, fungus, hormone-mediated signaling pathways, defense, osmotic stress, oxidative stress and water deprivation (TAIR). Glycosyl hydrolase family 1 (GH1) β-glucosidases in rice also responses to biotic and abiotic stress (Opassiri et al., 2006). When a plant is subjected to abiotic stress, the content of soluble sugars tends to increase. This elevation in soluble sugars has been associated with an enhancement in the plant's stress resistance. In

the case of soybeans under drought stress, the expression level of *Glyma.18G266900* was observed to decrease (Figure 5), This reduction in expression suggests a potential strategy employed by the plant to mitigate drought damage. It is conceivable that the plant decreases the expression of glycosyl hydrolase (*Glyma.18G266900*) to inhibit the hydrolysis of sugars, thus providing an additional protective mechanism against the adverse effects of drought stress.

In conclusion, the identified candidate genes are likely key contributors to soybean's response to drought stress. Further comprehensive research is essential to unravel the specific functions of these genes in soybean drought tolerance, providing insights into their molecular mechanisms. This deeper understanding holds the potential to inform strategies for enhancing abiotic stress resistance in soybean varieties through targeted genetic improvements.

Developing KASP markers

KASP (Kompetitive Allele-Specific PCR) marker technology holds significant importance and has wide applications in various fields, particularly in agriculture. Within agriculture, KASP markers are extensively used for multiple purposes, including germplasm resource identification, genetic relationship research, molecular marker-assisted breeding, genetic map construction and gene mapping. KASP has been applied to locate candidate genes for yield traits such as heading date, plant height, and thousand grain weight (Xiong et al., 2021; Xie et al., 2022); quality traits such as the color and shape of vegetables, fruits and other crops (Cheng et al., 2021; Shen et al., 2021); and genes of biotic and abiotic (Wang et al., 2018; Liu et al., 2020). In the current study, we developed two KASP markers S14_5147797 and S18_53902767 associated with drought tolerance (Figure 6). This achievement holds substantial application value for breeding drought-resistant soybean varieties. By using KASP markers, researchers and breeders can efficiently select and develop soybean varieties with improved drought tolerance, contributing to sustainable agriculture and food security.

The KASP markers developed in this study provide a valuable tool for soybean breeding programs aiming to enhance drought resistance.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding authors.

Author contributions

QJ: Data curation, Funding acquisition, Methodology, Writing – original draft, Writing – review & editing. MZ: Methodology, Software, Writing – review & editing. YX: Methodology, Software, Writing – review & editing. JW: Data curation, Methodology, Writing – review & editing. DX: Supervision, Writing – review & editing. HZ: Supervision, Writing – review & editing. XL: Supervision, Writing – review & editing. WZ: Methodology, Software, Supervision, Writing – review & editing. QW: Software, Supervision, Writing – review & editing. XS: Data curation, Investigation, Methodology, Supervision, Writing – review & editing. HC: Funding acquisition, Resources, Supervision, Visualization, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This study was supported by grants from Jiangsu Funding Program for Excellent Postdoctoral Talent (2023ZB647), Project funded by China Postdoctoral Science Foundation (2023M731401) and Zhongshan Biological Breeding Laboratory (ZSBBL).

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.2024.1352379/full#supplementary-material

References

Arya, H., Singh, M. B., and Bhalla, P. L. (2021). Towards developing drought-smart soybeans. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.750664

Bocian, A., Zwierzykowski, Z., Rapacz, M., Koczyk, G., Ciesiolka, D., and Kosmala, A. (2015). Metabolite profiling during cold acclimation of Lolium perenne genotypes distinct in the level of frost tolerance. *J. Appl. Genet.* 56, 439–449. doi: 10.1007/s13353-015-0293-6

Cheng, Z., Liu, Z., Xu, Y., Ma, L., Chen, J., Gou, J., et al. (2021). Fine mapping and identification of the candidate gene BFS for fruit shape in wax gourd (Benincasa hispida). *Theor. Appl. Genet.* 134, 3983–3995. doi: 10.1007/s00122-021-03942-8

Cotrim, M. F., Gava, R., Campos, C. N. S., De David, C. H. O., Reis, I. D. A., Teodoro, L. P. R., et al. (2021). Physiological performance of soybean genotypes grown under irrigated and rainfed conditions. J. Agron. *Crop Sci.* 207, 34–43. doi: 10.1111/jac.12448

Devi, J. M., Sinclair, T. R., Chen, P., and Carter, T. E. (2014). Evaluation of elite southern maturity soybean breeding lines for drought tolerant traits. *Agron. J.* 106, 1947–1954. doi: 10.2134/agronj14.0242

Du, L., Huang, X., Ding, L., Wang, Z., Tang, D., Chen, B., et al. (2023). TaERF87 and TaAKS1 synergistically regulate TaP5CS1/TaP5CR1-mediated proline biosynthesis to enhance drought tolerance in wheat. *New Phytol.* 237, 232–250. doi: 10.1111/nph.18549

Hacisalihoglu, G., Burton, A. L., Gustin, J. L., Eker, S., Asikli, S., Heybet, E. H., et al. (2018). Quantitative trait loci associatedwith soybean seed weight and composition under different phosphorus levels. *J. Integr. Plant Biol.* 60, 232–241. doi: 10.1111/jipb.12612

Hossain, M. S., Persicke, M., ElSayed, A. I., Kalinowski, J., and Dietz, K. J. (2017). Metabolite profiling at the cellular and subcellular level reveals metabolites associated with salinity tolerance in sugar beet. *J. Exp. Bot.* 68, 5961–5976. doi: 10.1093/jxb/erx388

Kaplan, F., Kopka, J., Haskell, D. W., Zhao, W., Schiller, K. C., Gatzke, N., et al. (2004). Exploring the temperature-stress metabolome of Arabidopsis. *Plant Physiol.* 136, 4159–4168. doi: 10.1104/pp.104.052142

Kishor, P. B. K., Suravajhala, R., Rajasheker, G., Marka, N., Shridhar, K. K., Dhulala, D., et al. (2020). Lysine, lysine-rich, serine, and serine-rich proteins: link between metabolism, development, and abiotic stress tolerance and the role of ncrnas in their regulation. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.546213

Knapp, S. J., Stroup, W. W., and Ross, W. M. (1985). Exact confidence intervals for heritability on a progeny mean basis. *Crop Sci.* 25, 192–194. doi: 10.2135/cropsci1985.0011183X002500010046x

Li, P. Y., Sun, Z. J., and Abulaiti, (2010). Evaluation of drought resistance of 29 accessions of elytrigria repens at seed germination stage under PEG-6000 stress. *Chinese Journal of Grassland* 32, 32–39. doi: CNKI:SUN:ZGCD.0.2010-01-006

Li, M., Guo, R., Jiao, Y., Jin, X., Zhang, H., and Shi, L. (2017). Comparison of salt tolerance in soja based on metabolomics of seedling roots. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.01101

Li, Z. (2021). *Population history and selection mode during soybean domestication* (Nangchang, China: Doctoral Dissertation of Nanchang University in China). doi: 10.27232/d.cnki.gnchu.2021.001153

Li, J. P., Xie, L., Han, D. Z., Gu, Y. Z., and Pi, Z. (2022). Evaluation of drought tolerance during germination of early-maturing soybean germplasm resources. *Heilongjiang. Agric. Sci.*, 1–6. doi: 10.13430/j.cnki.jpgr.20200330001

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, 639–649. doi: 10.1111/nph.12291

Liu, S., Bai, G., Lin, M., Luo, M., Zhang, D., Jin, F., et al. (2020). Identification of candidate chromosome region of Sbwm1 for Soil-borne wheat mosaic virus resistance in wheat. *Sci. Rep.* 10, 8119. doi: 10.1038/s41598-020-64993-3

Liu, Y., Gai, J. Y., Lv, H. N., Wang, Y. J., and Chen, S. Y. (2005). Identification of drought tolerant germplasm and in heritance and QTL mapping of related root traits in soybean (Glycine max(L.) Merr). *J. Genet. Genomics* 32, 855–863. doi: 10.1016/0008-6215(90)80036-3

Maughan, S. C., Pasternak, M., Cairns, N., Kiddle, G., Brach, T., Jarvis, R., et al. (2010). Plant homologs of the Plasmodium falciparum chloroquine-resistance transporter, PfCRT, are required for glutathione homeostasis and stress responses. *Proc. Natl. Acad. Sci. U.S.A.* 107, 2331–2336. doi: 10.1073/pnas.0913689107

Mavrodiev, E. V., Dervinis, C., Whitten, W. M., Gitzendanner, M. A., Kirst, M., Kim, S., et al. (2021). A new, simple, highly scalable, and efficient protocol for genomic DNA extraction from diverse plant taxa. *Appl. Plant Sci.* 9, e11413. doi: 10.1002/aps3.11413

Opassiri, R., Pomthong, B., Onkoksoong, T., Akiyama, T., Esen, A., and Ketudat Cairns, J. R. (2006). Analysis of rice glycosyl hydrolase family 1 and expression of Os4bglu12 β-glucosidase. *BMC Plant Biol.* 6, 33. doi: 10.1186/1471-2229-6-33

Ouyang, W., Chen, L., Ma, J., Liu, X., Chen, H., Yang, H., et al. (2022). Identification of quantitative trait locus and candidate genes for drought tolerance in a soybean recombinant inbred line population. *Int. J. Mol. Sci.* 23, 10828. doi: 10.3390/ijms231810828

Poudel, S., Vennam, R. R., Shrestha, A., Reddy, K. R., Wijewardane, N. K., Reddy, K. N., et al. (2023). Resilience of soybean cultivars to drought stress during flowering and early-seed setting stages. *Sci. Rep.* 13, 1277. doi: 10.1038/s41598-023-28354-0

Saini, R., and Nandi, A. K. (2022). TOPLESS in the regulation of plant immunity. *Plant Mol. Biol.* 109, 1–12. doi: 10.1007/s11103-022-01258-9

Saleem, A., Roldán-Ruiz, I., Aper, J., and Muylle, H. (2022). Genetic control of tolerance to drought stress in soybean. *BMC Plant Biol.* 22, 615. doi: 10.1186/s12870-022-03996-w

Sedivy, E. J., Wu, F., and Hanzawa, Y. (2017). Soybean domestication: Te origin, genetic architecture and molecular bases. *New. Phytol.* 214, 539–553. doi: 10.1111/nph.14418

Shen, J., Xu, X., Zhang, Y., Niu, X., and Shou, W. (2021). Genetic mapping and identification of the candidate genes for mottled rind in *Cucumis melo L. Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.769989

Singh, B., Khurana, P., Khurana, J. P., and Singh, P. (2018). Gene encoding vesicle-associated membrane protein-associated protein from Triticum aestivum (TaVAP) confers tolerance to drought stress. *Cell Stress Chaperones.* 23, 411–428. doi: 10.1007/s12192-017-0854-1

Stewart, G. R., and Larher, F. (1980). *Accumulation of amino acids and related compounds in relation to environmental stressin* (Berlin, Heidelberg: Springer Berlin Heidelberg: Academic Press), 609–635. doi: 10.1016/B978-0-12-675405-6.50023-1

Sun, B., Wang, Y., Yang, Q., Gao, H., Niu, H., Li, Y., et al. (2023). A high-resolution transcriptomic atlas depicting nitrogen fixation and nodule development in soybean. *J. Integr. Plant Biol.* 65, 1536–1552. doi: 10.1111/jipb.13495

Walla, A., Wilma van Esse, G., Kirschner, G. K., Guo, G., Brünje, A., Finkemeier, I., et al. (2020). An acyl-CoA N-Acyltransferase regulates meristem phase change and plant architecture in barley. *Plant Physiol.* 183, 1088–1109. doi: 10.1104/pp.20.00087

Wang, K., Bu, T., Cheng, Q., Dong, L., Su, T., Chen, Z., et al. (2021b). Two homologous LHY pairs negatively control soybean drought tolerance by repressing the abscisic acid responses. *New Phytol.* 229, 2660–2675. doi: 10.1111/nph.17019

Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 38, e164. doi: 10.1093/nar/gkq603

Wang, L. B., Liu, L. J., Pei, Y. F., Dong, S. K., Sun, C. S., Zu, W., et al. (2012). Drought resistant identification of soybean germplasm resources at bud stage. *J. Northeast Agric. Univ.* 43, 36–43. doi: 10.19720/j.cnki.issn.1005-9369.2012.01.007

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

Wang, H., Zhao, S., Mao, K., Dong, Q., Liang, B., Li, C., et al. (2018). Mapping QTLs for water-use efficiency reveals the potential candidate genes involved in regulating the trait in apple under drought stress. *BMC Plant Biol.* 18, 136. doi: 10.1186/s12870-018-1308-3

Waterworth, W. M., Bray, C. M., and West, C. E. (2015). The importance of safe guarding genome integrity in germination and seed longevity. *J. Exp. Bot.* 66, 3549–3558. doi: 10.1093/jxb/erv080

Xie, X., Li, S., Liu, H., Xu, Q., Tang, H., Mu, Y., et al. (2022). Identification and validation of a major QTL for kernel length in bread wheat based on two F3 biparental populations. *BMC Genomics* 23, 386. doi: 10.1186/s12864-022-08608-3

Xie, Z., Nolan, T. M., Jiang, H., and Yin, Y. (2019). AP2/ERF transcription factor regulatory networks in hormone and abiotic stress responses in Arabidopsis. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.00228

Xiong, H., Li, Y., Guo, H., Xie, Y., Zhao, L., Gu, J., et al. (2021). Genetic mapping by integration of 55K SNP array and KASP markers reveals candidate genes for important agronomic traits in hexaploid wheat. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.628478

Yamaguchi, N., Matsubara, S., Yoshimizu, K., Seki, M., Hamada, K., Kamitani, M., et al. (2021). H3K27me3 demethylases alter HSP22 and HSP17.6C expression in response to recurring heat in Arabidopsis. *Nat. Commun.* 12, 3480. doi: 10.1038/s41467-021-23766-w

Yang, J., Gao, M. X., Hu, H., Ding, X. M., Lin, H. W., Wang, L., et al. (2016). OsCLT1, a CRT-like transporter 1, is required for glutathione homeostasis and arsenic tolerance in rice. *New Phytol.* 211, 658–670. doi: 10.1111/nph.13908

Yang, W. M., Wang, M., Li, G. Q., and Du, W. J. (2013). Germination traits of soybean BIL population under PEG stress and assessment of their drought tolerance. *Chin. J. Oil Crop Sci.* 035, 564–571.

Zhang, H. Y., Jiao, B. C., and Li, G. Q. (2005). Study on selecting targets in drought-resistant breeding of soybean. *Soyb. Sci.*, 183–188.

Zhang, Z., Ma, J., Yang, X., Liu, Z., Liu, Y., Liu, X., et al. (2023). Natural allelic diversities of GmPrx16 confer drought tolerance in soybean. *Plant Biotechnol. J.* doi: 10.1111/pbi.14249

Zhang, W., Xu, W., Zhang, H., Liu, X., Cui, X., Li, S., et al. (2021). Comparative selective signature analysis and highresolution GWAS reveal a new candidate gene controlling seed weight in soybean. *Theor. Appl. Genet.* 134, 1329–1341. doi: 10.1007/s00122-021-03774-6

Zhao, Y., Cao, P., Cui, Y., Liu, D., Li, J., Zhao, Y., et al. (2021). Enhanced production of seed oil with improved fatty acid composition by overexpressing NAD^+ -dependent glycerol-3-phosphate dehydrogenase in soybean. *J. Integr. Plant Biol.* 63, 1036–1053. doi: 10.1111/jipb.13094

Zhao, X., Liu, Z., Li, H., Zhang, Y., Yu, L., Qi, X., et al. (2022). Identification of drought-tolerance genes in the germination stage of soybean. *Biol. (Basel).* 11, 1812. doi: 10.3390/biology11121812

Zhao, C., Liu, B., Piao, S., Wang, X., Lobell, D. B., Huang, Y., et al. (2017). Temperature increase reduces global yields of major crops in four independent estimates. *Proc. Natl. Acad. Sci. U.S.A.* 114, 9326–9331. doi: 10.1073/pnas.1701762114

Zia, R., Nawaz, M. S., Siddique, M. J., Hakim, S., and Imran, A. (2021). Plant survival under drought stress: implications, adaptive responses, and integrated rhizosphere management strategy for stress mitigation. *Microbiol. Res.* 242, 126626. doi: 10.1016/j.micres.2020.126626

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 27 March 2024
DOI 10.3389/fpls.2024.1369650

OPEN ACCESS

EDITED BY
Li Song,
Yangzhou University, China

REVIEWED BY
Chengfu Su,
Qingdao Agricultural University, China
Yumin Wang,
Jilin Academy of Agricultural Sciences (CAAS), China

*CORRESPONDENCE
Cunyi Yang
ycy@scau.edu.cn
Yingxiang Wang
yx_wang@fudan.edu.cn

†These authors have contributed equally to this work

RECEIVED 12 January 2024
ACCEPTED 07 March 2024
PUBLISHED 27 March 2024

CITATION
Liu G, Fang Y, Liu X, Jiang J, Ding G, Wang Y, Zhao X, Xu X, Liu M, Wang Y and Yang C (2024) Genome-wide association study and haplotype analysis reveal novel candidate genes for resistance to powdery mildew in soybean.
Front. Plant Sci. 15:1369650.
doi: 10.3389/fpls.2024.1369650

Genome-wide association study and haplotype analysis reveal novel candidate genes for resistance to powdery mildew in soybean

Guoqiang Liu[1,2†], Yuan Fang[3†], Xueling Liu[1,2†], Jiacan Jiang[1,2], Guangquan Ding[1,2], Yongzhen Wang[1,2], Xueqian Zhao[1,2], Xiaomin Xu[1,2], Mengshi Liu[1,2], Yingxiang Wang[3,4*] and Cunyi Yang[1,2*]

[1]Guangdong Provincial Key Laboratory of Plant Molecular Breeding, College of Agriculture, South China Agricultural University, Guangzhou, China, [2]Key Laboratory for Enhancing Resource Use Efficiency of Crops in South China, Ministry of Agriculture and Rural Affairs, South China Agricultural University, Guangzhou, China, [3]Guangdong Laboratory for Lingnan Modern Agriculture, Guangzhou, China, [4]Guangdong Provincial Key Laboratory of Protein Function and Regulation in Agricultural Organisms, College of Life Sciences, South China Agricultural University, Guangzhou, China

Powdery mildew disease (PMD) is caused by the obligate biotrophic fungus *Microsphaera diffusa* Cooke & Peck (*M. diffusa*) and results in significant yield losses in soybean (*Glycine max* (L.) Merr.) crops. By identifying disease-resistant genes and breeding soybean accessions with enhanced resistance, we can effectively mitigate the detrimental impact of PMD on soybeans. We analyzed PMD resistance in a diversity panel of 315 soybean accessions in two locations over 3 years, and candidate genes associated with PMD resistance were identified through genome-wide association studies (GWAS), haplotype analysis, qRT-PCR, and EMS mutant analysis. Based on the GWAS approach, we identified a region on chromosome 16 (Chr16) in which 21 genes form a gene cluster that is highly correlated with PMD resistance. In order to validate and refine these findings, we conducted haplotype analysis of 21 candidate genes and indicated there are single nucleotide polymorphisms (SNPs) and insertion-deletions (InDels) variations of *Glyma.16G214000, Glyma.16G214200, Glyma.16G215100* and *Glyma.16G215300* within the coding and promoter regions that exhibit a strong association with resistance against PMD. Subsequent structural analysis of candidate genes within this cluster revealed that in 315 accessions, the majority of accessions exhibited resistance to PMD when *Glyma.16G214300, Glyma.16G214800* and *Glyma.16G215000* were complete; however, they demonstrated susceptibility to PMD when these genes were incomplete. Quantitative real-time PCR assays (qRT-PCR) of possible candidate genes showed that 14 candidate genes (*Glyma.16G213700, Glyma.16G213800, Glyma.16G213900, Glyma.16G214000, Glyma.16G214200, Glyma.16G214300, Glyma.16G214500, Glyma.16G214585, Glyma.16G214669, Glyma.16G214700, Glyma.16G214800, Glyma.16G215000, Glyma.16G215100* and *Glyma.16G215300*) were involved in PMD resistance. Finally, we evaluated the PMD resistance of mutant lines from the Williams 82 EMS mutations library, which revealed that mutants of *Glyma.16G214000, Glyma.16G214200, Glyma.16G214300, Glyma.16G214800, Glyma.16G215000,*

Glyma.16G215100 and *Glyma.16G215300*, exhibited sensitivity to PMD. Combined with the analysis results of GWAS, haplotypes, qRT-PCR and mutants, the genes *Glyma.16G214000, Glyma.16G214200, Glyma.16G214300, Glyma.16G214800, Glyma.16G215000, Glyma.16G215100* and *Glyma.16G215300* were identified as highly correlated with PMD resistance. The candidate genes identified above are all NLR family genes, and these discoveries deepen our understanding of the molecular basis of PMD resistance in soybeans and will be useful for guiding breeding strategies.

KEYWORDS

GWAS, PMD, haplotypes, qRT-PCR, EMS mutations, NLR

1 Introduction

Soybean [*Glycine max* (L.) Merr.] is a leguminous crop that provides approximately 71% of plant-based protein and 29% of oil globally (Lin et al., 2022) and is an important source of animal and aquaculture feed (Adak and Kibritci, 2016). However, the proportion of soybean yield loss due to disease is increasing (Lin et al., 2022). Powdery mildew disease (PMD) is a common soybean disease caused by the fungus *Microsphaera diffusa* (*M. diffusa*) particularly in temperatures ranging from 15°C to 30°C. When the temperature is below 15°C or above 30°C, PMD infections will be reduced (Alves et al., 2009). It is easily detected on seeds, stems, leaves, and roots as white and powdery patches (Ramalingam et al., 2020) that result in defoliation, chlorosis veinal, necrosis, or mixtures of several symptoms (Grau, 2006). PMD causes yield reductions of up to 35% in susceptible soybean accessions (Dunleavy, 1980; Phillips, 1984) and is an epidemic disease in Australia (McTaggart et al., 2012), Canada (Takamatsu et al., 2002), Peru (Takamatsu et al., 2002), Puerto Rico (Takamatsu et al., 2002), Venezuela (Takamatsu et al., 2002), Brazil (Goncalves et al., 2002), Asia (Li M. W. et al., 2020), northeast India (Baiswar et al., 2016), and the United States (Grau, 2006). Despite the global importance of PMD as a soybean disease, the molecular basis of resistance and susceptibility to it remain largely uncharacterized.

Several PMD resistance loci in multiple soybean accessions have been mapped to the Chr16, including *Rmd_V97-3000* (Wang et al., 2013), *Rmd_PI243540* (Kang and Mian, 2010), *Rmd_PI567301B* (Jun et al., 2012), *Rmd_B3* (Jiang et al., 2019), *Rmd_B13* (Jiang et al., 2019), and *Rmd_ZH24* (Zhou et al., 2022). Mapping indicates they overlap or partially overlap each other suggesting that they could be tightly linked loci or one gene with different alleles. *Rmd_B13* was mapped to a genomic region containing 17 disease resistance (*R*)-like genes (Jiang et al., 2019), and *Rmd_ZH24* to an interval with 4 disease *R*-like genes (Zhou et al., 2022). Recently, the first soybean PMD resistance gene, *GmRmd1*, has been cloned through a combination of multiple methods, including a genome-wide association study from 467 soybean accessions, map-based cloning of 471 F_8 recombinant inbred lines derived from Guizao1 (susceptible) × B13 (resistant), and denovo assembly of the Guizao1 and B13 draft genomes using single-molecule long-read sequencing technology that can explore SVs in the *GmRmd1* region (Xian et al., 2022). By contrast, genes associated with PMD resistance have been extensively studied in other plants. In *Arabidopsis thaliana*, the *RESISTANCE TO POWDERY MILDEW 8.2* (*RPW8.2*) gene encodes phosphatase type 2C (PAPP2C), which negatively regulates salicylic acid (SA)-dependent basal defense against PMD (Wang et al., 2012). The barley *Mla* locus contains *Mla1* (Zhou et al., 2001), *Mla12* (Shen et al., 2003), and *Mla13* (Lu et al., 2016), which are nucleotide-binding and leucine-rich repeat (NLR) family proteins that recognize avirulence (AVR) proteins from the PMD fungus *Blumeria graminis* f. sp. *hordei*. In wheat, nearly 70 PMD resistance loci have been identified (Muller et al., 2022), but only a few genes have been cloned, including *Pm3* (Yahiaoui et al., 2004), *Pm8* (Hurni et al., 2013), *Pm17* (Singh et al., 2018), *Pm21* (Singh et al., 2018; Xing et al., 2018), *Pm60* (Zou et al., 2017), *Pm5* (Xie et al., 2020), and *Pm41* (Li M. et al., 2020). They all encode NLR family proteins.

Genome-wide association study (GWAS) is a genetic marker detection technique that has evolved into a pivotal method for investigating the genetics of intricate diseases (Burghardt et al., 2017). Compared to conventional linkage analysis, GWAS can significantly enhance the precision and accuracy of marker–phenotype associations (Yano et al., 2016; Cortes et al., 2021). Currently, GWAS has been successfully employed in elucidating the genetic basis underlying soybean PMD resistance. In 2020, the first GWAS analysis of soybean PMD resistance was used in gene mapping research, and this study obtained 30,510 high-quality SNP loci from 331 soybean accessions for association analysis resulting in the identification of *Glyma.16g210800*, *Glyma.16g211000*, and *Glyma.16g211400* as important candidate genes for disease resistance (Hu, 2021). In addition, Xian utilized GWAS to classify the genetic polymorphisms associated with PMD resistance in soybeans, identified a single region associated with

PMD resistance based on 2,176,969 SNPs in 467 soybean accessions, and finally identified *GmRmd1* as the PMD resistance gene in this region (Xian et al., 2022). Another study identified seven SNPs significantly associated with resistance to PMD through GWAS, and combined with differential expression levels, three candidate genes for *Rmd_ZDD00359* were determined: *Glyma.16G210800*, *Glyma.16G212300*, and *Glyma.16G213900* (Sang et al., 2023). In recent years, this method has been utilized to identify numerous QTLs and genes governing crucial disease-resistant traits in soybean. For instance, through GWAS analysis, a total of 36,976 SNP markers associated with resistance to soybean cyst nematode (SCN) *Heterodera glycines* (HG) Ichinohe 0 type and *Heterodera glycines* Ichinohe 1.2.3.5.7 type were identified across a diverse panel of 440 soybean accessions, and a total of 19 associated signals were detected and significantly correlated with the resistance of two types of HG (Han et al., 2015; Cortes et al., 2021). A GWAS analysis was performed on a total of 330 soybean accessions, resulting in the identification of 25,179 SNPs and the discovery of eight genomic regions significantly associated with resistance to soybean white mold (Boudhrioua et al., 2020). Utilizing a whole-genome association mapping approach, 800 soybean accessions were employed to identify genomic regions associated with resistance to phytophthora root and stem rot, and 16 SNP markers located on chromosomes 3, 13, and 19 exhibited a significant correlation with resistance of soybeans to this disease (Schneider et al., 2016). Therefore, GWAS holds significant potential for enhancing disease resistance in soybean breeding.

The objective of this article is to investigate the role of NLR gene clusters in conferring soybean resistance against PMD, thereby establishing a fundamental understanding of plant disease resistance mechanisms and providing valuable insights for breeding PMD-resistant soybean accessions.

2 Materials and methods

2.1 Plant materials and evaluation of PMD resistance

Williams 82 (W82) and Huaxia 3 (HX3) were provided by South China Agricultural University. The 315 soybean accessions used for this study were obtained from soybean research institutions in China (Supplementary Table 1). Accessions were planted in December in field sites located in Guangzhou and Hainan, and these were used for disease evaluation and genetic analysis.

In December 2017, accessions were sown in the field of Hainan Experimental Station, China. Disease resistance was evaluated in February 2018, and the data were recorded as Y2018. In December 2018 and December 2019, accessions were grown in the field at South China Agricultural University, Guangzhou. PMD resistance was evaluated in March 2019 and March 2020, and the data were recorded as Y2019 and Y2020, respectively. All plants were scored and counted for PMD resistance (Supplementary Table 1). All plants were challenged with *M. diffusa* as described previously (Jiang et al., 2019). Each plant was inoculated with spores of *M. diffusa* at stage V1 by brushing with PMD-infected leaves of susceptible plants maintained in the greenhouse, as described by Kang and Mian (2010). Three weeks after inoculation, PMD incidence was evaluated for the 315 soybean accessions. Accessions were scored for PMD resistance using a scale. A plant with no PMD colonies on any leaf was regarded as a resistant (R) line and scored "L0," while a plant with one or more PMD colonies on its leaves was rated as a susceptible (S) line, depending on disease severity, from L1 grade to L5 (Supplementary Table 1) (Li et al., 2016).

2.2 DNA extraction and analysis

The sodium dodecyl sulfonate (SDS) method was used to extract DNA from GWAS populations from young unfolded trifoliolate leaves at stage V2 (second trifoliolate) (Perry, 2004). A minimum of 30 ng/μl of DNA, with an OD260/280 of 1.8–2.0, was used for resequencing, and greater than 2 μg of DNA was used to generate small fragment libraries for pair-end sequencing.

2.3 Genome mapping and detection of SNPs and InDels

A total of 315 soybean accessions were selected for whole-genome re-sequencing using the Illumina HiSeq X Ten platform with 150-bp pair-end reads. The average sequencing depth of all samples is 15×. A detailed sequencing information is shown in Supplementary Table 2. All the sequencing in this study was done by Annoroad Gene Technology Co. Ltd. (Zhejiang, China).

Version 4 of Williams 82 genome in the Phytozome database (https://phytozome-next.jgi.doe.gov/info/Gmax_Wm82_a4_v1) was used as the reference for data analysis. Read quality control to remove low-quality reads and adapter sequences was done using Fastp (version 0.23.2) (Chen et al., 2018) with default parameters. Clean reads from each sample were aligned to the reference genome using BWA-MEM 2 (release 2.2.1) (Vasimuddin et al., 2019), and then, mapped reads were sorted using SAMtools (version 1.15.1) (Li et al., 2009). GATK4 packages (release 4.2.6.1) (Mckenna et al., 2010) were employed for variation detection and haplotype analysis. Duplicate-read marking was performed using MarkDuplicates. Variants from each sample were called using HaplotypeCaller to generate files in gVCF format. The gVCF files were merged using CombineGVCFs generating a VCF file composed of emit-all-sites information. The VCF file was filtered using the recommended GATK parameters "QUAL < 30.0 || QD < 2.0 || MQ < 40.0 || FS > 60.0 || SOR > 3.0 || MQRankSum < −12.5 || ReadPosRankSum < −8.0" and "QUAL < 30.0 || QD < 2.0 || FS > 200.0 || SOR > 10.0 || ReadPosRankSum < -20.0 || MQ < 40.0 || MQRankSum < −12.5" for SNPs and InDels, respectively. Apart from this, SNPs were filtered using the following criteria: removing SNPs as missing greater than 15% and MAF lower than 0.5% among 315 accessions.

2.4 Detecting significant SNPs

TASSEL (version 5.2.8.1) (Bradbury et al., 2007) was used to calculate the association of the PMD resistance phenotype with genetic polymorphisms in 315 soybean accessions by applying a mixed linear model (MLM). In this study, admixture software was employed to estimate the population structure matrix (Q). When K = 11, the CV error was the minimum value at 0.54741. TASSEL5 software was used to calculate the kinship matrix (K) with the parameter-method Centered_IBS. As covariates, Q and K were used to control the population in the MLM analysis. As part of the TASSEL analysis, the phenotypic variation explained (PVE) was also calculated using the mixed liner model. The CMplot package (Yin et al., 2021) was used for visualizing the Manhattan p and QQ plots. The Manhattan plot shows the p-values of significant SNPs associated with PMD resistance. The QQ plot shows the relationship between theory and expectation. SnpEff (Cingolani et al., 2012) was used for gene annotation of heterotopic points according to the physical location and structure of genes throughout the genome.

2.5 Analysis of linkage disequilibrium and haplotype blocks

LD analysis measures the nonrandom association of pairs of SNPs. TASSEL (version 5.2.8.1) (Bradbury et al., 2007) software calculates the square (R2) of the allele frequency correlation of SNP. The core SNP set derived after suitable filtering was used for calculating linkage disequilibrium (LD) patterns and the structure of haplotype blocks of SNPs by LDBlockShow (Dong et al., 2020). Data sets (r2 and D′) of pairwise LD measurement calculations showed LD levels of SNPs.

2.6 Haplotype analysis of candidate genes

The genotype code types and physical position in the genome of the candidate locus were obtained from genotype-calling VCF format files. Unsupervised clustering was applied to three genotype codes consisting of reference homozygote, alternative homozygote, and missing by Cluster3.0 software (Hoon et al., 2004).

2.7 Evaluation of the expression patterns of 21 candidate genes

V1 stage W82 and HX3 plants were infected with an *M. diffusa* spore suspension containing 1×10^5 cfu/ml and maintained in a growth chamber at 23°C, 75% relative humidity, with a 16-h light/8-h dark photoperiod to characterize the expression of 21 candidate genes. The experiments included three replicates. Leaves were sampled at 0, 6, 12, 24, 48, and 72 h after inoculation and kept at −80°C. Total RNA was extracted using the HiPure Plant RNA Mini Kit (Magen, Guangzhou, Guangdong, China), and 1 mg of total RNA was reverse transcribed to produce first-strand cDNA using the *Evo M-MLV* RT Kit with gDNA Clean for qRT-PCR (Accurate Biology, Guangzhou, Guangdong, China). Candidate genes in the target region were predicted in SoyBase using the Wm82.a4.v1 reference genome. Quantitative real-time polymerase chain reaction (qRT-PCR) was performed to obtain the expression profiles of candidate genes using primers designed with Primer Premier 5.0 (Premier, Vancouver, Canada). The housekeeping gene *Actin* was used as a control. The specific primers for 21 candidate genes are listed in Supplementary Table 6. The qRT-PCR was performed with a CFX96 Real-Time PCR Detection System (Bio-Rad, Hercules, CA, USA) using SYBR® Green Premix *Pro Taq* HS qPCR Kit II (Accurate Biology, Guangzhou, Guangdong, China). All reactions were performed in 20-μl volumes containing 1 μl of cDNA as a template. The thermal cycling conditions consisted of 94°C for 2 min, followed by 40 cycles of 95°C for 10 s, 59°C for 30 s, and 72°C for 30 s. Three independent biological repeats were used. The qRT-PCR data were evaluated using the $2^{-\triangle\triangle CT}$ method (Livak and Schmittgen, 2001).

2.8 The PMD resistance evaluation of mutant lines from Williams 82 EMS mutation library

The information of mutant lines from the Williams 82 EMS-induced library was extracted from this website: http://isoybean.org/ (Supplementary Table 7). Seeds for mutants were obtained from Song's laboratory (Zhang et al., 2022). These mutant lines were grown in the greenhouse at South China Agricultural University, Guangzhou. The mutant lines were subjected to two generations of selfing to obtain homozygous lines of the corresponding candidate genes (Supplementary Table 7). Each mutant plant was inoculated with spores of *M. diffusa* at stage V1. After 2 weeks of inoculation, the *M. diffusa* of the mutant leaves was observed and photographed, and these leaves were stored at −20°C for DNA extraction and sequencing.

2.9 Phylogenetic analysis

Candidate gene protein sequence from the reference genome annotation Wm82.a4.v1 version was used to blast the RefSeq Select proteins database (https://www.ncbi.nlm.nih.gov/refseq/refseq_select/). The top 100 sequences with the highest similarity were selected for complete multiple sequence alignment. The phylogenetic tree was constructed using the NJ method in MEGA X (Kumar et al., 2018) with 1,000 bootstrap replicates.

3 Results

3.1 Evaluation of soybean germplasm for resistance to PMD

To study PMD resistance in soybeans, we collected 315 soybean accessions from around the world and evaluated their resistance to PMD during 3 years in Hainan province (P.R. China) and an

experimental field at South China Agricultural University (SCAU). We divided our screening results into five levels (L1–L5) of PMD susceptible according to the previous studies (Kang and Mian, 2010) and defined PMD resistant (level 0; L0). Representative phenotypes of highly PMD susceptible (L5), medium sensitivity (L3), and resistant (L0) are shown in Figure 1A, and the PMD resistance (L0) or susceptible level (L1–L5) of each accession is detailed in Supplementary Table 1. During the spring of 2018, we identified 196 resistant and 97 susceptible lines at the Hainan Experimental Station (Figure 1B). During the spring of 2019, we identified 207 resistant and susceptible 98 lines at the experimental station at South China Agricultural University (SCAU) (Figure 1B). During the spring of 2020, we identified 205 resistant and 110 susceptible accessions at the SCAU experimental station(Figure 1B). Pooling these results across years and locations, we found that 206 (66.35%) of the 315 accessions are resistant to PMD, and 109 (34.60%) are susceptible. These results indicate that soybean accessions are a rich resource for PMD resistance (Supplementary Table 1).

3.2 SNPs/InDels genotyping and GWAS for PMD resistance

We performed whole-genome re-sequencing of 315 soybean accessions from China, the United States, Brazil, Australia, and Africa (Supplementary Table 2), with an average 15× depth of sequencing. We mapped 92.65% of the reads to the soybean reference genome (Wm82.a4.v1 version) and used GATK to call 4,900,642 high-quality SNPs and 702,316 InDels (<40 bp) at bi-allelic loci. A distribution analysis of SNPs and InDels across the 20 chromosomes of the soybean genomes shows a high density near the ends of chromosomes 3, 6, 15, 16, and 18 (Supplementary Figure 1). Each chromosome has an average of 245,032 SNPs and 35,116 InDels.

We used the SNPs and InDels from the diversity collection of 315 accessions to conduct a GWAS with a mixed linear model approach to identify variant sites that are significantly associated with PMD resistance (Figures 1C–F). A total of 56 SNPs surpassed the genome-wide significance threshold clustered at the end of the long arm of chromosome 16 (Figure 1G). The SNP locus with the highest –log10 (p-value) of 19.14 is at nucleotide position 37,418,263 on Chr16 and accounts for 33.38% (R2) of the phenotypic variance. This GWAS peak is flanked by two paralogous R genes (*Glyma.16G214200* and *Glyma.16G214300*). Two other genes (*Glyma.16G213700* and *Glyma.16G214641*) overlap several SNPs in peaks with the second highest significance (Figure 1G). We defined 13 genes in this region as candidate PMD resistance genes (Supplementary Table 3). Furthermore, this region encompasses 8 NLR genes potentially associated with resistance to PMD, thereby yielding a total of 21 candidate genes for subsequent comprehensive analysis (Supplementary Table 3).

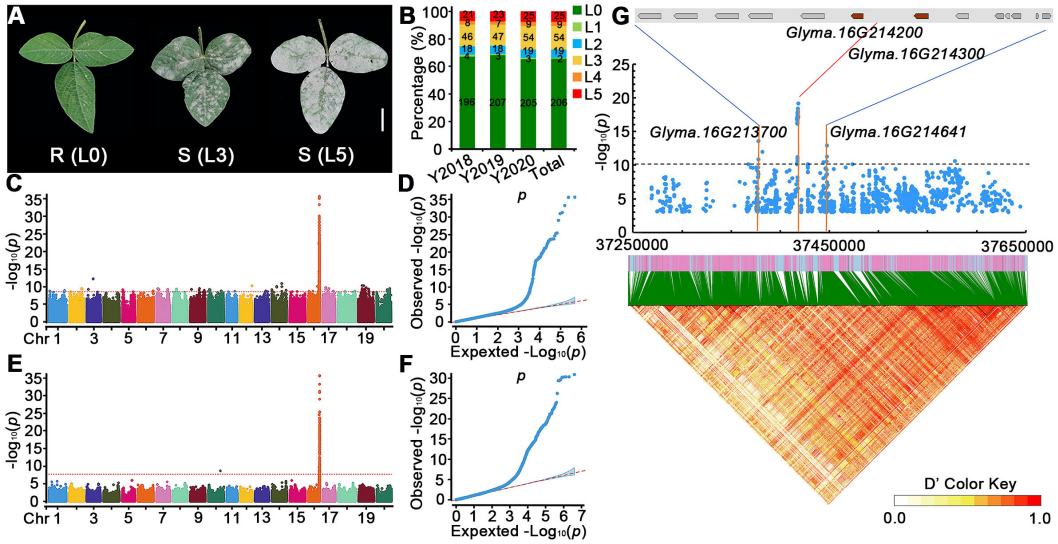


FIGURE 1
Genome-wide association analysis of PMD resistance genes in the 315 soybean accessions. **(A)** Phenotypic traits of soybean accessions after 20 days of *M. diffusa* infection. R, resistant; S, susceptible. Scale bar = 2 cm. **(B)** Percentages of PMD-resistant levels of 315 accessions in different regions and years. L0 represents resistant to PMD, and L1–L5 represent susceptible levels to PMD increasingly in **(A, B)**. Y2018: Hainan, February 2018; Y2019: SCAU field, March 2019; Y2020: SCAU field, March 2020. **(C–F)** GWAS scan for PMD using re-sequencing data of 315 accessions grown in spring from different regions and years; **(C, D)** represent Manhattan plots of significant SNPs associated with PMD and quantile–quantile plots for PMD, respectively. **(E, F)** represent Manhattan plots of significant InDels associated with PMD and quantile–quantile plots for PMD, respectively. The GWAS results are presented as negative log10 p-values against position on each of the 20 chromosomes. Horizontal red dashed lines indicate the genome-wide significant threshold. **(G)** Local Manhattan plot (top) and linkage disequilibrium heatmap (bottom) surrounding the peak on Chr16. Two genes of brown color (*Glyma.16G214200* and *Glyma.16G214300*) are flanked on both sides by the most significant SNPs (top panel). A local Manhattan plot and linkage disequilibrium heatmap of that 400-kb region are displayed in the bottom panel. The orange lines indicate the candidate region for the peak. The red line indicates the highest peak value. The blue plot indicates the nucleotide variation of the candidate genes. Horizontal black dashed lines indicate the genome-wide significant threshold. The colors in the figure from yellow to red represent the D′ value from low degree to high, which is the standardized disequilibrium coefficient between a pair of alleles.

We used LD Blockshow to detect haplotype blocks in the 400-kb region containing the most significant SNP markers (Figure 1G). There is strong linkage disequilibrium across the region. The most significant SNP out of 60 significant SNPs in the block (located at Chr16: 37,416,562–37,418,263) (Supplementary Figure 2) is upstream of *Glyma.16G214200* and downstream of *Glyma.16G214300* suggesting that PMD resistance may be associated with variants in DNA regulatory elements of *Glyma.16g214200* or *Glyma.16g214300*.

3.3 Haplotype analysis of the candidate genes

We analyzed SNPs and InDels at the 21-gene (*Glyma.16G213700*–*Glyma.16G215400*) locus in the diversity pane of 315 soybean accessions. Unsupervised clustering was used to cluster all SNP variant types based on three genotype codes: reference homozygote, SNP/InDel mutation (Alter), and missing (Figure 2). There is a mutation caused by a base substitution at position 746 bp in the coding region of the *Glyma.16G214000*, which is associated with 71.56% (78/109, out of 109 susceptible accessions, 78 contained the mutation) of susceptible accessions (Figure 2). SNP mutations in the upstream promoter region of *Glyma.16G214200* from −1,534 to −1,978 bp are associated with 95% (104/109) of susceptible accessions (Figure 3). Furthermore, several NLR candidate genes with relatively high peak values are located between *Glyma.16G214641* and *Glyma.16G215400* suggesting their potential involvement in PMD resistance. Haplotype analysis of these genes identified an SNP at position 72 in the coding region of the *Glyma.16G215100* gene, and this mutation was present in 55% (60/109) of susceptible accessions (Figure 4). There is a 3-bp deletion at 226 bp in the coding region of *Glyma.16G215300*, and 91% (99/109) of susceptible accessions are associated with this mutation (Figure 5). The haplotypes of other candidate genes are depicted in Supplementary Figures 3–17. In summary, SNP/InDel mutations in the coding and promoter regions of the above genes are closely related to their disease resistance, and these genes may be candidate genes for resistance to PMD.

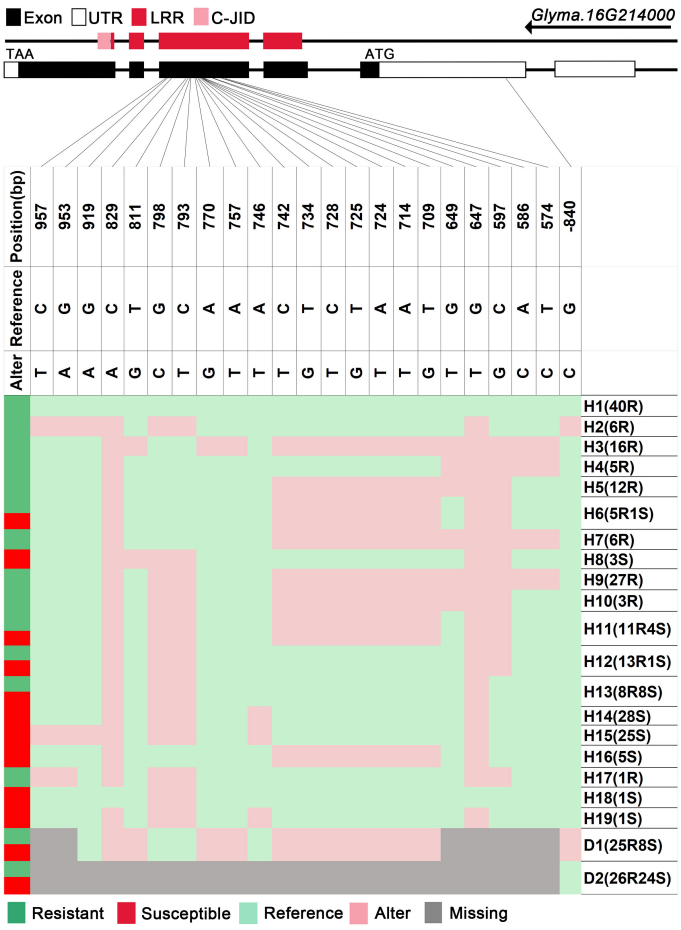


FIGURE 2
Haplotype analysis of *Glyma.16G214000*. H, haplotype; R, resistant; S, susceptible; D1, deletion type 1; H6(5R1S), five accessions in H6 are resistant, and one is susceptible; LRR, leucine-rich repeat domin; C-JID, C-terminal jelly roll/Ig-like domain.

10.3389/fpls.2024.1369650

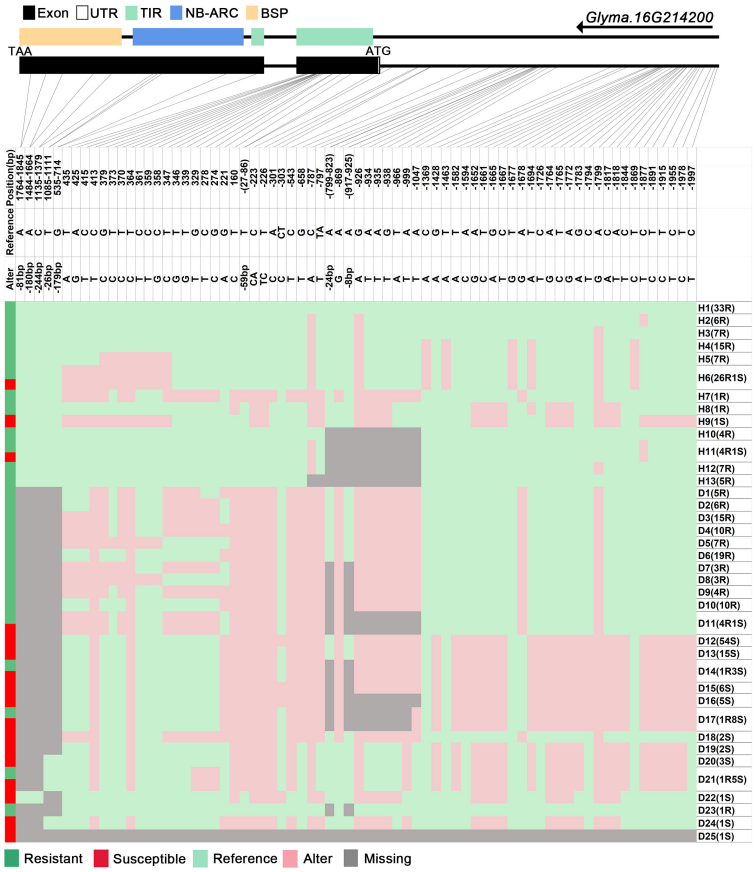


FIGURE 3
Haplotype analysis of *Glyma.16G214200*. H, haplotype; R, resistant; S, susceptible; D1, deletion type 1; H6(26R1S), 26 accessions in H6 are resistant, and 1 is susceptible; TIR, Toll/interleukin-1 receptor; NBS, nucleotide-binding site; BSP, basic secretory proteins.

3.4 Haplotype analysis of other candidate genes in 11 special accessions at the *GmRmd1* locus

The *GmRmd1* (*Glyma.16G214300*) gene has been previously confirmed as a resistant gene against PMD in various studies (Xian et al., 2022). In this study, the majority of accessions possessing intact *GmRmd1* coding regions demonstrate PMD resistance, whereas those lacking such regions exhibit susceptibility to PMD (Figure 6). These findings underscore the critical importance of maintaining an intact coding region for *GmRmd1* in conferring soybean resistance against accessions. However, the integrity of the *GmRmd1* coding region was not consistent with the PMD phenotype in 11 of the 315 accessions in which three of the resistant accessions exhibited incomplete *GmRmd1*; concurrently, eight of the susceptible accessions displayed intact *GmRmd1* (Supplementary Table 4). This suggests the presence of additional genetic factors contributing to resistance against PMD. Subsequently, through the analysis of haplotypes in other candidate genes (*Glyma.16G213700*–*Glyma.16G215400*) within these 11 accessions, and by integrating high-quality haplotype data (Supplementary Tables 4, 5), it was observed that among the eight susceptible accessions, haplotypes (D3 of *Glyma.16G213700*), (H18 and H19 of *Glyma.16G213800*), (H15 and H16 of *Glyma.16G214000*), (D18 of *Glyma.16G214200*), (H11 of *Glyma.16G214500*), (H8, H19, and D7 of *Glyma.16G214529*), (H9 of *Glyma.16G214557*), (H41 of *Glyma.16G214585*), (H37, H40, and D1 of *Glyma.16G214613*), (H32, H50, H51, and D1 of *Glyma.16G214641*), (H12 and H18 of *Glyma.16G215100*), and (H8 of *Glyma.16G215300*) were found to be susceptible in other accessions. Among the three disease-resistant accessions, haplotypes (H4 of *Glyma.16G214529*), (H29 and H36 of *Glyma.16G214585*), (H8 of *Glyma.16G214613*), (H6 of *Glyma.16G215100*), and (H2 of *Glyma.16G215300*) exhibited resistance to diseases in other accessions. In conclusion, it is plausible that *Glyma.16G213700*, *Glyma.16G213800*, *Glyma.16G214000*, *Glyma.16G214200*, *Glyma.16G214500*, *Glyma.16G214557*, *Glyma.16G214585*, *Glyma.16G214613*, *Glyma.16G214641*, *Glyma.16G215100*, and *Glyma.16G215300* may represent additional genes exhibiting resistance against PMD.

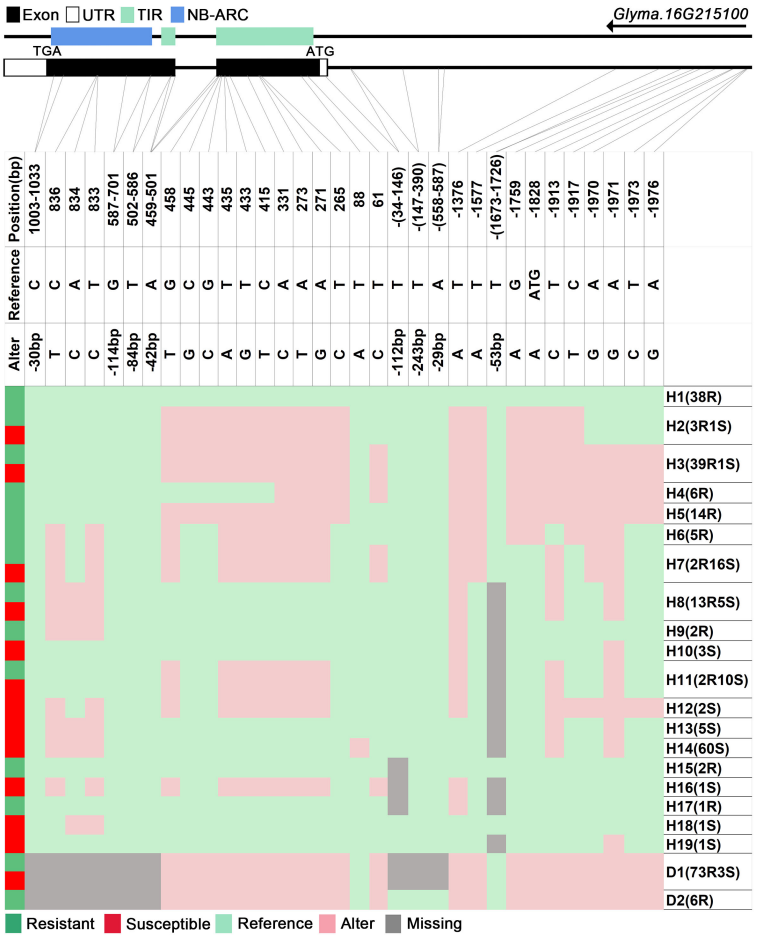


FIGURE 4
Haplotype analysis of *Glyma.16G215100*. H, haplotype; R, resistant; S, susceptible; D1, deletion type 1; H2(3R1S), three accessions in H2 are resistant, and one is susceptible; TIR, Toll/interleukin-1 receptor; NBS, nucleotide-binding site.

3.5 The assessment of the integrity of coding regions in candidate genes

Through structural analysis of the coding regions of candidate genes, it was observed that among 206 disease-resistant accessions, a total of 203 exhibited intact *Glyma.16G214300*; conversely, out of 109 susceptible accessions, only 101 displayed incomplete *Glyma.16G214300* (Figure 7). Furthermore, within the cohort of disease-resistant accessions, a significant majority (193 out of 206) possessed intact *Glyma.16G214800*; in contrast, among the susceptible group (consisting of 109 samples), only 82 showed incomplete *Glyma.16G214800* (Figure 7). Similarly, within the population of disease-resistant accessions (206 in total), a substantial proportion (185) demonstrated intact *Glyma.16G215000*; however, among the susceptible counterparts (comprising 109 samples), merely 68 exhibited incomplete *Glyma.16G215000* (Figure 7). This observation suggests a strong association between the integrity of the coding region for *Glyma.16G214300*, *Glyma.16G214800*, and *Glyma.16G215000* and their disease resistance phenotypes, and the resistance gene *Glyma.16G214300* has been confirmed against PMD (Xian et al., 2022) suggesting that genes *Glyma.16G214800* and *Glyma.16G215000* are highly likely to confer resistance to PMD as well. It is plausible that these genes form a complex network involved in conferring disease resistance by synergistically exerting protective effects against PMD.

3.6 Expression patterns for candidate genes

To investigate the expression of 21 candidate genes in response to *M. diffusa* infection, qRT-PCR was utilized to examine the expression patterns of each gene between Williams 82 (W82, resistance to PMD) and Huaxia 3 (HX3, susceptible to PMD) (Figure 8). Among these genes, the expression levels of *Glyma.16G213900* and *Glyma.16G214000* were predominantly observed in W82, while they were scarcely detected in HX3. In comparison to HX3, the expression levels of *Glyma.16G214100*, *Glyma.16G214300*, *Glyma.16G214500*, *Glyma.16G214585*, *Glyma.16G214800*, *Glyma.16G215100*, and *Glyma.16G215300*

10.3389/fpls.2024.1369650

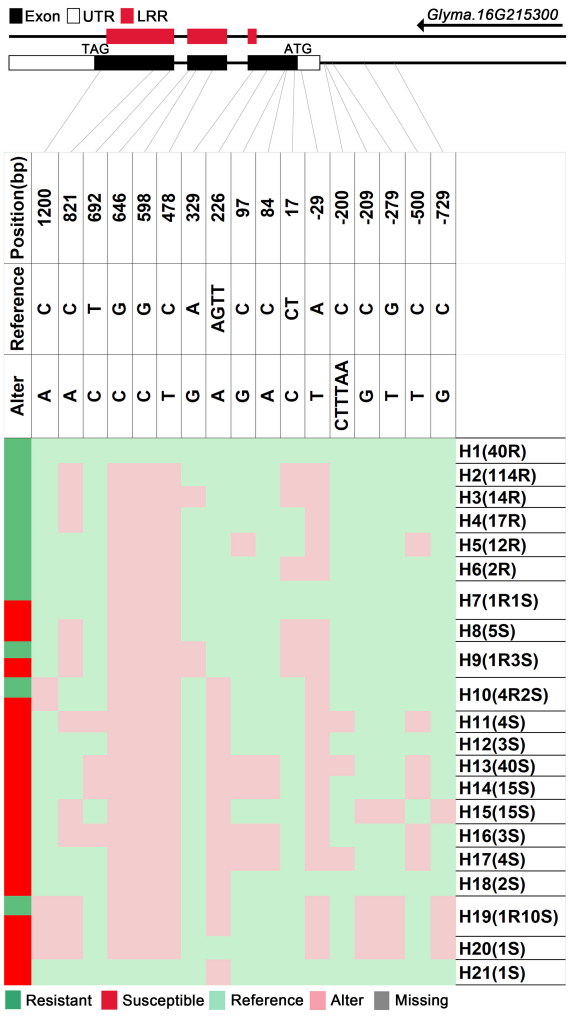


FIGURE 5
Haplotype analysis of *Glyma.16G215300*. H, haplotype; R, resistant; S, susceptible; D1, deletion type 1; H7(1R1S), one accession in H7 is resistant, and one is susceptible; LRR, leucine-rich repeat domin.

were upregulated highly in W82. Conversely, the expression levels of *Glyma.16G214529*, *Glyma.16G214557*, *Glyma.16G214641*, and *Glyma.16G215000* showed lower expression levels in W82 compared with HX3. The expression levels of *Glyma.16G214613* and *Glyma.16G214669* exhibited similar patterns between W82 and HX3. However, *Glyma.16G213700*, *Glyma.16G213800*, and *Glyma.16G214700* were expressed in HX3 but not detected in W82 suggesting their potential involvement in negative regulation of PMD resistance.

Additionally, six genes (*Glyma.16G213900*, *Glyma.16G214000*, *Glyma.16G214500*, *Glyma.16G214300*, *Glyma.16G214585*, and *Glyma.16G214800*) were induced by *M. diffusa* in W82 but showed minimal or low expression in HX3, and the expression levels of *Glyma.16G213900*, *Glyma.16G214000*, and *Glyma.16G214500* gradually increased in W82 after *M. diffusa* infection, reaching their peak after 48 h and subsequently decreasing. Correspondingly, the expression levels of the *Glyma.16G214300*, *Glyma.16G214585*, and *Glyma.16G214800* genes peaked after 24 h followed by a gradual decline. These results indicated that the above six genes could positively regulate PMD resistance. In addition, five genes (*Glyma.16G214200*, *Glyma.16G214669*, *Glyma.16G215000*, *Glyma.16G215100*, and *Glyma.16G215300*) were induced to express in W82 after being infected by *M. diffusa*. Among them, the expression level of *Glyma.16G214200* reached its peak at 24 h, while that of *Glyma.16G215000* peaked at 48 h. Similarly, the highest expression levels for *Glyma.16G214669*, *Glyma.16G215100*, and *Glyma.16G215300* were observed after 72 h. However, these five genes exhibited distinct expression patterns in HX3 at each time point, suggesting a potential coordinated role of these genes in the regulation of PMD, necessitating their collaboration with other resistance genes for effective control of *M. diffusa* infection. To summarize, a total of 14 genes may be involved in the regulation of PMD resistance, with three genes specifically implicated in negative regulation.

3.7 Williams 82 mutant lines confirm that multiple genes are involved in soybean PMD resistance

Mutant lines ranging from *3700* (*Glyma.16G213700*) to *5400* (*Glyma.16G215400*), derived from Williams 82 induced with EMS, were obtained from Song's laboratory (Zhang et al., 2022), and the specific mutation sites and homozygosity of each line are depicted in Figures 9B–J and Supplementary Table 7. Based on the phenotypic characteristics of mutant lines, it can be inferred that *rmd1-1*, *rmd1-2*, *rmd1-3*, *4000*, *4200*, *4800*, *5000*, *5100*, and *5300* lines exhibit susceptibility to PMD, while the mutant lines of other candidate genes, such as *3800* and *5400*, are resistant to PMD (Figure 9A). Among them, the gene *Glyma.16G214300* in the mutant lines *rmd-1*, *rmd-2*, and *rmd-3* has been identified as a known resistance gene against PMD (Xian et al., 2022). However, *Glyma.16G214300* was not mutated in lines *3800*, *4000*, *4200*, *4800*, *5000*, *5100*, *5300*, and *5800* (Supplementary Figures 18-22). This suggests that apart from *Glyma.16G214300*, *Glyma.16G214000*, *Glyma.16G214200*, *Glyma.16G214800*, *Glyma.16G215000*, *Glyma.16G215100*, and *Glyma.16G215300* also significantly contribute to PMD resistance.

4 Discussion

PMD, caused by the fungus *M. diffusa*, has resulted in substantial yield losses for soybeans. The identification of resistant genes and the breeding of resistant cultivars under optimal environmental conditions can effectively mitigate the detrimental impact imposed by soybean PMD. Previous studies used populations to identify soybean PMD resistance loci, which are concentrated at the end of Chr16 (35,968,531–37,733,538 bp; use Wm82.a2.v1 reference genome). These loci are *Rmd_V97-3000* at

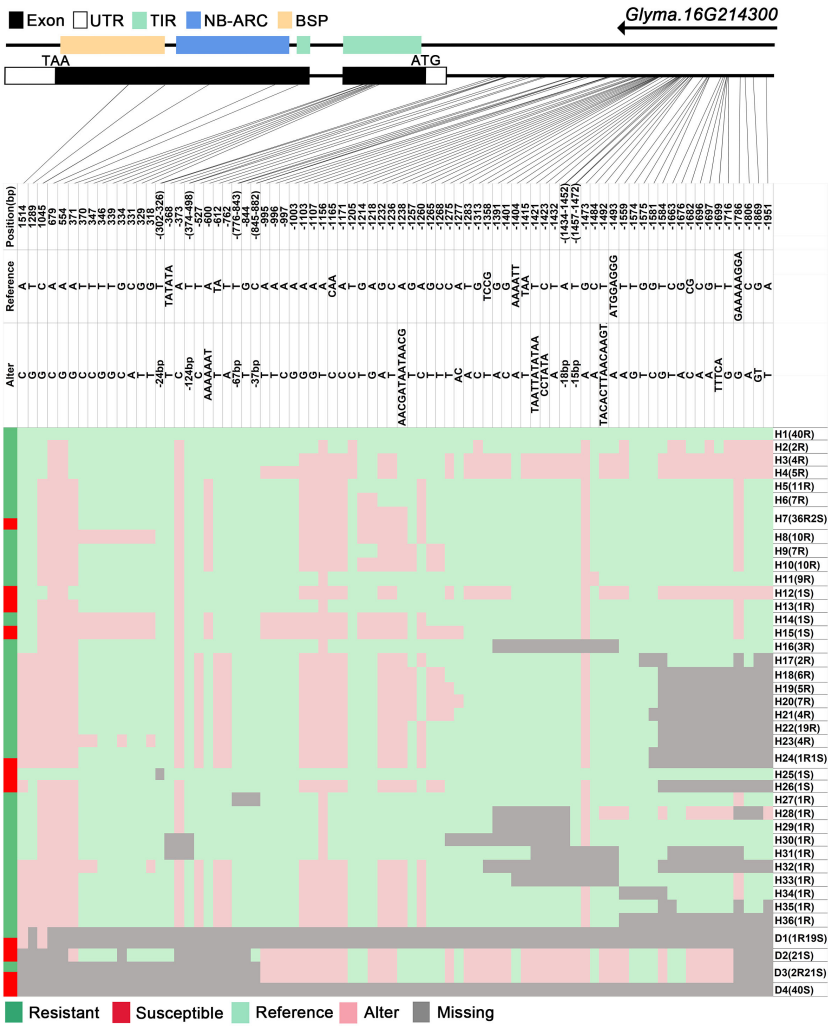


FIGURE 6
|. Haplotype analysis of *Glyma.16G214300*. H, haplotype; R, resistant; S, susceptible; D1, deletion type 1; H7(36R2S), 36 accessions in H6 are resistant, and 2 are susceptible; TIR, Toll/interleukin-1 receptor; NBS, nucleotide-binding site; BSP, basic secretory proteins.

34,035,391–37,631,694 (Wang et al., 2013), *Rmd_PI24540* at 34,258,523–36,750,257 (Kang and Mian, 2010), *PMD_PI567301B* (5000–6700) at 37,249,583–37,370,175 bp (Jun et al., 2012), *Rmd_B3* at 36,221,397–37,631,694 (Jiang et al., 2019), *Rmd_B13* at 37,102,014-37,290,074 (Jiang et al., 2019), *Rmd_ZH24* at 37,202,495–37,235,283 (Zhou et al., 2022), and *Rmd_ZDD00359* at 37,011,583–37,234,234 (Sang et al., 2023). They completely or partially overlap one another and may be an R-gene cluster or a single locus. In these loci, *Rmd_V97-3000*, *Rmd_B3*, *Rmd_B13*, *Rmd_ZH24*, and *Rmd_ZDD00359* all harbor the *GmRmd1* gene. Additionally, it was later confirmed that the resistance gene in *Rmd_B13* is indeed *GmRmd1*. Besides these findings, the loci *Rmd_PI24540* and *PMD_PI567301B* do not possess the *GmRmd1* gene indicating the presence of other yet-to-be-identified genes responsible for PMD resistance. The *PMD_PI567301B* locus harbors 16 genes, including *Glyma.16G215000*, *Glyma.16G215100*, *Glyma.16G215300*, and *Glyma.16g215400* as NLR genes, suggesting that one of these four genes may confer resistance to soybean PMD caused by *PMD_PI567301B*. Hu (2021) analyzed 331 soybean accessions through GWAS and identified three candidate genes (*Glyma.16g210800*, *Glyma.16g211000*, and *Glyma.16g211400*) treated as PMD resistance genes. Another study also identified three candidate genes (*Glyma.16G210800*, *Glyma.16G212300*, and *Glyma.16G213900*) that were considered PMD-resistant genes by GWAS analysis (Sang et al., 2023).The results indicate that different accessions resistant to PMD may carry various resistance genes.

In this study, we used GWAS to identify a region on the end of Chr16 associated with PMD resistance (Figure 1E). This region contains 21 genes, with the exception of *Glyma.16G214613* and *Glyma.16G214641*, as the remaining 19 genes all belong to the NLR gene family (Supplementary Figure 23). Among these, the gene *Glyma.16G214200* encodes a protein that demonstrates similarity to *GmRmd1* with a BSP (basic secretory protein, peptidase of plants and bacteria) domain. The BSP domain has similarity to the

FIGURE 7
Integrity analysis of candidate gene-coding regions: 3700: *Glyma.16g213700*; 3800: *Glyma.16g213800*, and so on. Levels of susceptibility: Disease levels 1 to 5.

defense-related proteins *WCI-5* from wheat (Görlach et al., 1996), *NtPRp27* from tobacco (Okushima et al., 2000), and *StPRp27* from potato (Shi et al., 2012). *GmRmd1* and *Glyma.16G214200* are homologs of *WCI-5/NtPRp27*, with a shared homology of 84% (Supplementary Figure 24), indicating that *Glyma.16G214200* may play an important role in resistance to PMD. Recently, experimental evidence has demonstrated that the truncated protein TIR-NBS of *SRC7* (*GmRmd1*) exhibits robust resistance against SMV and TMV (Yan et al., 2022). Moreover, partial resistance is also observed in the truncated protein TIR-NBS of *SRC8* (*Glyma.16G214200*) suggesting that the BSP domain is not dispensable for antiviral activity (Yan et al., 2022). Through structural domain and evolutionary analysis of candidate proteins (Supplementary Figures 23, 24), it was discovered that *Glyma.16G215100* solely contains the TIR and NBS domains without the LRR or BSP domains at its C-terminus. Additionally, *Glyma.16G215100* shares significant similarity with truncated protein TN from *Glyma.16G214200* and *Glyma.16G214300* displaying strong resistance to both soybean mosaic virus (SMV) and tobacco mosaic virus (TMV) (Yan et al., 2022). These findings imply a potentially crucial role of *Glyma.16G215100* in PMD resistance.

Calcium ion-related compounds also play a crucial role in plant immune defense with many calcium-binding proteins containing EF-hand-type calcium-binding domains (Ikura et al., 2002; Nelson et al., 2002). The SRC4 (Glyma.16G214800) proteins possess a Toll/interleukin-1 receptor (TIR) domain at the N-terminus and an EFh (EF-hand) domain at the C-terminus (Supplementary Figure 23), and proteins harboring these domains are implicated in conferring resistance against SMV and TMV (Yan et al., 2022) suggesting that *Glyma.16G214800* may also contribute to PMD resistance. The genes *Glyma.16G214000* and *Glyma.16g215000* are all NLR family genes that possess the C-JID domain (Supplementary Figure 23), and this domain is responsible for substrate recognition by binding to effector proteins of pathogens and plays a crucial role in initiating TIR-NLR receptor signaling. The RPS4 protein in Arabidopsis, which contains the C-JID (or posterior LRR) structure, collaborates with RRS1 to recognize effectors and subsequently exert disease resistance (Ma et al., 2020). The C-terminus of *GmRmd1* harbors a BSP domain while lacking the LRR domain responsible for direct pathogen recognition. Proteins possessing the C-JID domain, such as *Glyma.16G214000* and *Glyma.16g215000*, may interact with *GmRmd1* to collectively perceive effector factors and confer disease resistance effects.

Examining differential gene expression patterns has been regarded as a promising approach for gaining a deeper biological understanding of GWAS signals (Emilsson et al., 2008). In this study, we identified 21 candidate genes associated with PMD resistance through GWAS and haplotype analysis. Among them, three genes (*Glyma.16G213700*, *Glyma.16G213800*, and *Glyma.16G214700*), which are expressed only in HX3 but not in W82, may potentially compete with resistance genes for binding to other resistance genes during the regulation of PMD resistance, thereby exerting a negative regulatory effect on PMD resistance. In addition, 11 genes (*Glyma.16G213900*, *Glyma.16G214000*, *Glyma.16G214200*, *Glyma.16G214300*, *Glyma.16G214500*, *Glyma.16G214585*, *Glyma.16G214669*, *Glyma.16G214800*, *Glyma.16G215000*, *Glyma.16G215100*, and *Glyma.16G215300*) were induced by *M. diffusa* in W82, while 6 genes (*Glyma.16G213900*, *Glyma.16G214000*, *Glyma.16G214500*, *Glyma.16G214300*, *Glyma.16G214585*, and *Glyma.16G214800*) exhibited either repressed or constrained levels of expression in HX3 indicating that these 6 genes play important roles in PMD. However, the other five genes (*Glyma.16G214200*, *Glyma.16G214669*, *Glyma.16G215000*, *Glyma.16G215100*, and *Glyma.16G215300*) showed different expression patterns at different time points in HX3 indicating that these five genes may need to work together with other genes to act on *M. diffusa*. Previous studies have demonstrated that genes exhibiting differential expression patterns between accessions are frequently directly or indirectly associated with susceptibility or resistance outcomes. Conversely, genes displaying distinct expression dynamics over time may represent a general plant response to pathogen infection without necessarily conferring increased resistance (Calla et al., 2009). Therefore, the aforementioned 11 genes can be regarded as robust candidate genes for conferring resistance against PMD. Combined with haplotype, qRT-PCR, and

FIGURE 8
Relative expression levels of 21 candidate genes in W82 (resistant) and HX3 (susceptible). qRT-PCR analysis of the expression of 21 candidate genes at 0, 6, 12, 24, 48, and 72 h after an *M. diffusa* spore suspension containing 1 × 10^5 cfu/ml treatment. The data are shown as mean ± SD.

mutant analysis data, it was ultimately determined that seven genes (*Glyma.16G214000*, *Glyma.16G214200*, *Glyma.16G214300*, *Glyma.16G214800*, *Glyma.16G215000*, *Glyma.16G215100*, and *Glyma.16G215300*) were identified as being important in PMD defense.

In this study, numerous novel candidate genes for PMD resistance have been identified. However, their mode of action remains uncertain, whether they function independently or synergistically with the *GmRmd1*. It is plausible that these genes gradually enhance resistance to PMD through functional superposition. Drawing inspiration from the strategy employed in generating PMD-resistant and high-yielding *Tamlo R32* wheat mutants (Li et al., 2022), genome editing emerges as an appealing approach for progressively eliminating or introducing disease-resistant genes in both resistant and susceptible soybeans. In conclusion, this study identified several resistance genes except *GmRmd1* that was closely associated with PMD resistance, and these results provide important genetic resources for breeding scientists to develop PMD resistance accessions.

5 Conclusion

Several important pathogens, including PMD, occur in soybeans and cause significant yield reductions globally. In this

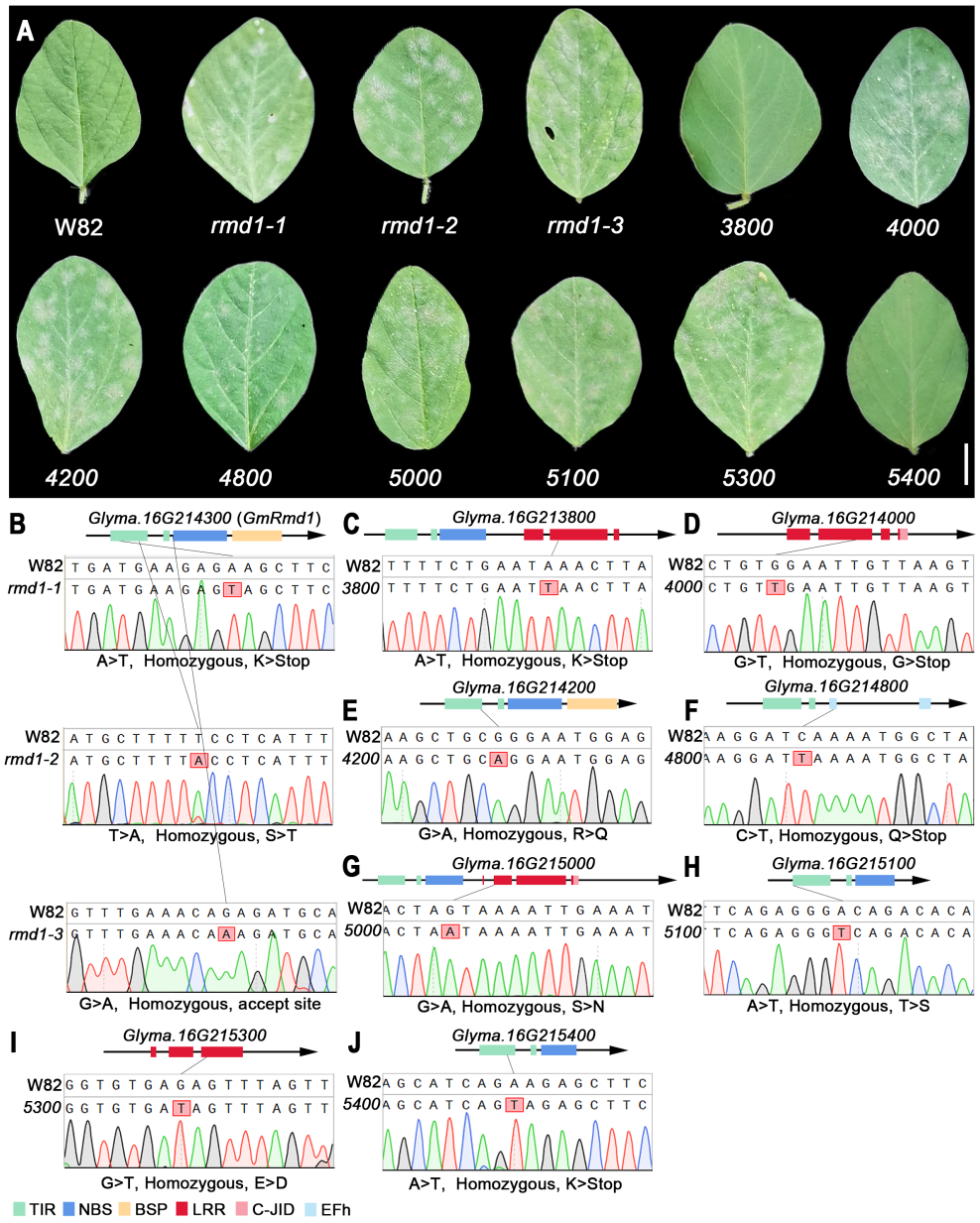


FIGURE 9
PMD resistance evaluation of candidate gene mutant lines. **(A)** Mutant phenotype of W82. W82: Williams 82, wild-type; *rmd1-1*, *rmd1-2*, *rmd1-3*: different mutants of *Glyma.16G214300*; *4000*, mutants of *Glyma.16G214000*, an so on. Scar bar: 1 cm. **(B–J)** Sequencing analysis of W82 mutant. **(B)** A > T/homozygous/K > Stop: Base A mutates into T/homozygous mutant/amino acid K mutates into codon terminator. **(C–J)** Consistent with the analysis method in **(B)**. TIR, Toll/interleukin-1 receptor; NBS, nucleotide-binding site; BSP, basic secretory proteins; LRR, leucine-rich repeat domin; C-JID, C-terminal jelly roll/Ig-like domain; EFh, EF-hand domain.

study, a genome-wide association study (GWAS) identified SNPs and InDels significantly associated with PMD resistance in a cluster of disease-resistant genes located at the distal end of chromosome 16. Haplotype, qRT-PCR, and mutant analysis revealed that candidate genes for resistance against PMD include *Glyma.16g214000*, *Glyma.16g214200*, *Glyma.16g214300*, *Glyma.16g214800*, *Glyma.16g215000*, *Glyma.16g215100*, and *Glyma.16g215300*. These findings establish a robust genetic basis for further elucidating the mechanisms underlying PMD resistance and facilitating breeding efforts toward developing resistant accessions.

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

GL: Writing – review & editing, Investigation, Formal analysis, Data curation. YF: Writing – review & editing, Validation, Formal

analysis. XL: Writing – review & editing, Investigation, Data curation. JJ: Writing – review & editing, Investigation, Data curation. GD: Writing – review & editing, Investigation. YZW: Writing – review & editing, Data curation. XZ: Writing – review & editing, Data curation. XX: Writing – review & editing, Data curation. ML: Writing – review & editing, Investigation, Data curation. YXW: Writing – review & editing, Writing – original draft, Supervision, Data curation. CY: Writing – review & editing, Writing – original draft, Supervision, Resources, Funding acquisition, Conceptualization.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the open competition program of top 10 critical priorities of the Agricultural Science and Technology Innovation for the 14th Five-Year Plan of Guangdong Province (2022SDZG05), the Major Program of Guangdong Basic and Applied Research (2019B030302006), the National Natural Science Foundation of China (Grant Nos. 32172061 and 32000403), and Guangdong Laboratory of Lingnan Modern Agriculture.

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.2024.1369650/full#supplementary-material

References

Adak, M., and Kibritci, M. (2016). Effect of nitrogen and phosphorus levels on nodulationand yield components in faba bean Vicia faba L. *Legume. Res.* 39, 991–994. doi: 10.18805/lr.v0iOF.3773

Alves, M. D. C., Pozza, E. A., Costa, J. C. B., Ferreira, J. B., and Araujo, D. V. D. (2009). Effects of temperature and leaf wetness period in powdery mildew *Microsphaera diffusa Cke. & Pk.* intensity in soybean [*Glycine max (L.) Merr.*] cultivars. *Cienc. Agrotec.* 33, 1926–1930. doi: 10.1590/S1413-70542009000700039

Baiswar, P., Chandra, S., and Ngachan, S. (2016). Molecular evidence confirms presence of anamorph of erysiphe diffusa on soybean (*glycine max*) in northeast India. *Australas. Plant Dis. Notes* 11, 25. doi: 10.1007/s13314-016-0213-6

Boudhrioua, C., Bastien, M., Torkamaneh, D., and Belzile, F. (2020). Genome-wide association mapping of Sclerotinia Sclerotiorum resistance in soybean using whole-genome resequencing data. *BMC Plant Biol.* 20, 195. doi: 10.1186/s12870-020-02401-8

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., Edward, S. B., et al. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23, 2633–2635. doi: 10.1093/bioinformatics/btm308

Burghardt, L. T., Young, N. D., and Tiffin, P. (2017). A guide to genome-wide association mapping in plants. *Curr. Protoc. Plant Biol.* 2, 22–38. doi: 10.1002/cppb.20041

Calla, B., Vuong, T., Radwan, O., Hartman, G. L., and Clough, S. J. (2009). Gene expression profiling soybean stem tissue early response to sclerotinia sclerotiorumand in silico mapping in relation to resistance markers. *Plant Genome-US.* 2, 149–166. doi: 10.3835/plantgenome2008.02.0008

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). Fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884–i890. doi: 10.1093/bioinformatics/bty560

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., et al. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly* 6, 80–92. doi: 10.4161/fly.19695

Cortes, L. T., Zhang, Z., and Yu, J. (2021). Status and prospects of genome-wide association studies in plants. *Plant Genome.* 14, e20077. doi: 10.1002/tpg2.20077

Dong, S. S., He, W. M., Ji, J. J., Zhang, C., and Yang, T. L. (2020). LDBlockShow: a fast and convenient tool for visualizing linkage disequilibrium and haplotype blocks based on variant call format files. *Briefings Bioinf.* 22, bbaa227. doi: 10.1093/bib/bbaa227

Dunleavy, J. M. (1980). Yield losses in soybeans induced by powdery mildew. *Plant Disease.* 64, 291–291. doi: 10.1094/pd-64-291

Emilsson, V., Thorleifsson, G., Zhang, B., Leonardson, A. S., Zink, F., Zhu, J., et al. (2008). Genetics of gene expression and its effect on disease. *Nature* 452, 423–428. doi: 10.1038/nature06758

Goncalves, E. C. P., Di Mauro, A. O., and Centurion, M. A. P. D. C. (2002). Genetics of resistance to powdery mildew (*Microsphaera diffusa*) in Brazilian soybean populations. *Gen. Mol. Biol.* 25, 339–342. doi: 10.1590/s1415-47572002000300015

Görlach, J., Volrath, S., Knauf-Beiter, G., Hengy, G., Beckhove, U., Kogel, K. H., et al. (1996). Benzothiadiazole, a novel class of inducers of systemic acquired resistance, activates gene expression and disease resistance in wheat. *Plant Cell.* 8, 629–643. doi: 10.2307/3870340

Grau, C. (2006) Powdery mildew of soybean. Available online at: http://fyi.uwex.edu/fieldcroppathology/files/2010/12/powdery_mildew_06.pdf.

Han, Y., Zhao, X., Cao, G., Wang, Y., Li, Y., Liu, D., et al. (2015). Genetic characteristics of soybean resistance to HG type 0 and HG type 1.2.3.5.7 of the cyst nematode analyzed by genome-wide association mapping. *BMC Genom.* 16, 598. doi: 10.1186/s12864-015-1800-1

Hoon, M. J. L. D., Imoto, S., Nolan, J., and Miyano, S. (2004). Open source clustering software. *Bioinformatics* 20, 1453–1454. doi: 10.1093/bioinformatics/bth078

Hu, L. (2021). Identification of resistance to powdery mildew in soybean resources and mapping of resistance gene associations. Jilin Agricultural University, Jilin. dissertation/master's thesis. doi: 10.27163/d.cnki.gjlnu.2020.000270

Hurni, S., Brunner, S., Buchmann, G., Herren, G., and Keller, B. (2013). Rye *Pm8* and wheat *Pm3* are orthologous genes and show evolutionary conservation of resistance function against powdery mildew. *Plant J.* 2013, 76. doi: 10.1111/tpj.1234

Ikura, M., Osawa, M., and Ames, J. B. (2002). The role of calcium-binding proteins in the control of transcription: structure to function. *BioEssays* 24, 625–636. doi: 10.1002/bies.10105

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

Jun, T. H., Mian, M. A. R., Kang, S. T., and Michel, A. P. (2012). Genetic mapping of the powdery mildew resistance gene in soybean PI 567301B. *Theor. Appl. Genet.* 125, 1159–1168. doi: 10.1007/s00122-012-1902-y

Kang, S.-T., and Mian, M. R. J. G. (2010). Genetic map of the powdery mildew resistance gene in soybean PI 243540. *Genome* 53, 400–405. doi: 10.1139/g10-015

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096

Li, M., Dong, L., Li, B., Wang, Z., Xie, J., Qiu, D., et al. (2020). A CNL protein in wild emmer wheat confers powdery mildew resistance. *New Phytol.* 228, 1027–1037. doi: 10.1111/nph.16761

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352

Li, S., Lin, D., Zhang, Y., Deng, M., Chen, Y., Lv, B., et al. (2022). Genome-edited powdery mildew resistance in wheat without growth penalties. *Nature* 602, 455–460. doi: 10.1038/s41586-022-04395-9

Li, M., Liu, N. X., Yue, Y. L., Jiang, B. Z., Li, X. P., and Nian, H. (2016). Preliminary screening for resistant soybean cultivars to powdery mildew in southern China. *Soybean Sci.* 35, 209–221. doi: 10.11861/j.issn.1000-9841.2016.02.0209

Li, M. W., Wang, Z., Jiang, B., Kaga, A., Wong, F. L., Zhang, G. H., et al. (2020). Impacts of genomic research on soybean improvement in East Asia.Theor. *Appl. Genet.* 133, 1655–1678. doi: 10.1007/s00122-019-03462-6

Lin, F., Chhapekar, S. S., Vieira, C. C., Da, S. M. P., Rojas, A., Lee, D., et al. (2022). Breeding for disease resistance in soybean: a global perspective. *Theor. Appl. Genet.* 135, 3773–3872. doi: 10.1007/s00122-022-04101-3

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2– ΔΔCT method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262

Lu, X., Kracher, B., Saur, I. M. L., Bauer, S., Ellwood, S. R., Wise, R., et al. (2016). Allelic barley MLA immune receptors recognize sequence-unrelated avirulence effectors of the powdery mildew pathogen. *PNAS* 113, E6486. doi: 10.1073/pnas.1612947113

Ma, S., Lapin, D., Liu, L., Sun, Y., Song, W., Zhang, X., et al. (2020). Direct pathogen-induced assembly of an NLR immune receptor complex to form holoenzyme. *Science* 370, eabe3069. doi: 10.1126/science.abe3069

Mckenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303. doi: 10.1101/gr.107524.110

McTaggart, A., Ryley, M., and Shivas, R. G. (2012). First report of the powdery mildew *Erysiphe diffusa* on soybean in Australia. *Australas. Plant Dis.* 7, 127–129. doi: 10.1007/s13314-012-0065-7

Muller, M. C., Kunz, L., Schudel, S., Lawson, A. W., Kammerecker, S., Wyler, M., et al. (2022). Ancient variation of the *AvrPm17* gene in powdery mildew limits the effectiveness of the introgressed rye *Pm17* resistance gene in wheat. *PNAS* 119, e2108808119. doi: 10.1073/pans.2108808119

Nelson, M. R., Thulin, E., Fagan, P. A., Forsé, S.., and Chazin, W. J. L. (2002). The EF-hand domain: A globally cooperative structural unit. *Protein Sci.* 11, 198–205. doi: 10.1110/ps.33302

Okushima, Y., Koizumi, N., Kusano, T., and Sano, H. (2000). Secreted proteins of tobacco cultured BY2 cells: identification of a new member of pathogenesis-related proteins. *Plant Mol. Biol.* 42, 479–488. doi: 10.1023/A:1006393326985

Perry, D. J. (2004). Identification of Canadian durum wheat varieties using a single PCR. *Theor. Appl. Genet.* 109, 55–61. doi: 10.1007/s00122-004-1597-9

Phillips, D. V. (1984). Stability of *Microsphaera diffusa* and the effect of powdery mildew on yield of soybean. *Plant Disease.* 68, 953–956. doi: 10.1094/PD-68-953

Ramalingam, J., Alagarasan, G., Savitha, P., Lydia, K., Pothiraj, G., Vijayakumar, E., et al. (2020). Improved host-plant resistance to Phytophthora rot and powdery mildew in soybean (*Glycine max (L.) Merr.*). *Sci. Rep.* 10, 13928. doi: 10.1038/s41598-020-70702-x

Sang, Y., Zhao, H., Liu, X., Yuan, C., Qi, G., Li, Y., et al. (2023). Genome-wide association study of powdery mildew resistance in cultivated soybean from Northeast China. *Front. Plant Sci.* 14. doi: 10.3389/fpls.2023.1268706

Schneider, R., Rolling, W., Song, Q., Cregan, P., Dorrance, A., and McHale, L. (2016). Genome-wide association mapping of partial resistance to *Phytophthora Sojae* in soybean plant introductions from the Republic of Korea. *BMC Genom.* 17, 607. doi: 10.1186/s12864-016-2918-5

Shen, Q. H., Zhou, F., Bieri, S., Haizel, T., and Schulze-Lefert, P. (2003). Recognition specificity and *RAR1/SGT1* dependence in barley *Mla* disease resistance genes to the powdery mildew fungus. *Plant Cell.* 15, 732–744. doi: 10.1105/tpc.009258

Shi, X., Tian, Z., Liu, J., Vossen, E. A. G. V. D., and Xie, C. (2012). A potato pathogenesis-related protein gene, *StPRp27*, contributes to race-nonspecific resistance against Phytophthora infestans. *Mol. Biol. Rep.* 39, 1909–1916. doi: 10.1007/s11033-011-0937-5

Singh, S. P., Hurni, S., Ruinelli, M., Brunner, S., Sanchez-Martin, J., Krukowski, P., et al. (2018). Evolutionary divergence of the rye *Pm17* and *Pm8* resistance genes reveals ancient diversity. *Plant Mol. Biol.* 98, 249–260. doi: 10.1007/s11103-018-0780-3

Takamatsu, S., Taguchi, Y., Shin, H. D., Paksiri, U., Limkaisang, S., Binh, N. T., et al. (2002). Two *Erysiphe* species associated with recent outbreak of soybean powdery mildew: results of molecular phylogenetic analysis based on nuclear rDNA sequences. *Mycoscience* 43, 333–341. doi: 10.1007/S102670200049

Vasimuddin, M., Misra, S., Li, H., and Aluru, S. (2019). "Efficient architecture-aware acceleration of BWA-MEM for multicore systems," in *Proceedings of the 2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS)*, (Piscataway, NJ: IEEE), 314–324.

Wang, W. M., Ma, X. F., Zhang, Y., Luo, M. C., Wang, G. L., Bellizzi, M., et al. (2012). PAPP2C interacts with the atypical disease resistance protein RPW8.2 and negatively regulates salicylic acid-dependent defense responses in Arabidopsis. *Mol. Plant* 5, 1125–1137. doi: 10.1093/mp/sss008

Wang, Y., Shi, A., Zhang, B., and Chen, P. (2013). Mapping powdery mildew resistance gene in V97-3000 soybean. *Plant Breeding.* 132, 625–629. doi: 10.1111/pbr.12072

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, 100418. doi: 10.1016/j.xplc.2022.100418

Xie, J., Guo, G., Wang, Y., Hu, T., Wang, L., Li, J., et al. (2020). A rare single nucleotide variant in *Pm5e* confers powdery mildew resistance in common wheat. *New Phytol.* 228, 1011–1026. doi: 10.1111/nph.16762

Xing, L., Hu, P., Liu, J., Witek, K., Zhou, S., Xu, J., et al. (2018). *Pm21* from haynaldia villosa encodes a CC-NBS-LRR protein conferring powdery mildew resistance in wheat. *Mol. Plant* 11, 874–878. doi: 10.1016/j.molp.2018.02.013

Yahiaoui, N., Srichumpa, P., Dudler, R., and Keller, B. (2004). Genome analysis at different ploidy levels allows cloning of the powdery mildew resistance gene *Pm3b* from hexaploid wheat. *Plant J.* 37, 528–538. doi: 10.1046/j.1365-313X.2003.01977.x

Yan, T., Zhou, Z. K., Wang, R., Bao, D. R., Li, S. S., Li, A. G., et al. (2022). A cluster of atypical resistance genes in soybean confers broad-spectrum antiviral activity. *Plant Physiol.* 188, 1277–1293. doi: 10.1093/plphys/kiab507

Yano, K., Yamamoto, E., Aya, K., Takeuchi, H., Lo, P. C., Hu, L., et al. (2016). Genome-wide association study using whole-genome sequencing rapidly identifies new genes influencing agronomic traits in rice. *Nat. Genet.* 48, 927–934. doi: 10.1038/ng.3596

Yin, L., Zhang, H., Tang, Z., Xu, J., Yin, D., Zhang, Z., et al. (2021). rMVP: A memory-efficient, visualization-enhanced, and parallel-accelerated tool for genome-wide association study. *Genomics. Proteomics Bioinf.* 19, 619–628. doi: 10.1016/j.gpb.2020.10.007

Zhang, M., Zhang, X., Jiang, X., Qiu, L., Jia, G., Wang, L., et al. (2022). iSoybean: A database for the mutational fingerprints of soybean. *Plant Biotechnol. J.* 20, 1435–1437. doi: 10.1111/pbi.13844

Zhou, Q., Jiang, B., Cheng, Y., Ma, Q., Xia, Q., Jiang, Z., et al. (2022). Fine mapping of an adult-plant resistance gene to powdery mildew in soybean cultivar Zhonghuang 24. *Crop J.* 10, 8. doi: 10.1016/j.cj.2021.12.003

Zhou, F., Kurth, J., Wei, F., Elliott, C., Valè, G., Yahiaoui, N., et al. (2001). Cell-autonomous expression of barley *Mla1* confers race-specific resistance to the powdery mildew fungus *via* a Rar1-independent signaling pathway. *Plant Cell.* 13, 337–350. doi: 10.1105/tpc.13.2.337

Zou, S., Wang, H., Li, Y., Kong, Z., and Tang, D. (2017). The NB-LRR gene Pm60 confers powdery mildew resistance in wheat. *New Phytol.* 218, 298–309. doi: 10.1111/nph.14964

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 15 April 2024
DOI 10.3389/fpls.2024.1387954

OPEN ACCESS

EDITED BY
Zhenbin Hu,
Agricultural Research Service (USDA), United States

REVIEWED BY
Hengyou Zhang,
Chinese Academy of Sciences (CAS), China
Fuan Niu,
Shanghai Academy of Agricultural Sciences, China

*CORRESPONDENCE
Donghe Xu
xudh@jircas.affrc.go.jp

†These authors have contributed equally to this work

RECEIVED 19 February 2024
ACCEPTED 29 March 2024
PUBLISHED 15 April 2024

CITATION
Kumawat G, Cao D, Park C and Xu D (2024) *C-terminally encoded peptide*-like genes are associated with the development of primary root at *qRL16.1* in soybean. *Front. Plant Sci.* 15:1387954. doi: 10.3389/fpls.2024.1387954

C-terminally encoded peptide-like genes are associated with the development of primary root at *qRL16.1* in soybean

Giriraj Kumawat[1,2†], Dong Cao[1,3†], Cheolwoo Park[1] and Donghe Xu[1*]

[1]Japan International Research Center for Agricultural Sciences, Tsukuba, Ibaraki, Japan, [2]Crop Improvement Section, ICAR-Indian Institute of Soybean Research, Indore, Madhya Pradesh, India, [3]Key Laboratory of Biology and Genetic Improvement of Oil Crops, Ministry of Agriculture and Rural Affairs, Oil Crops Research Institute, Chinese Academy of Agricultural Sciences, Wuhan, Hubei, China

Root architecture traits are belowground traits that harness moisture and nutrients from the soil and are equally important to above-ground traits in crop improvement. In soybean, the root length locus *qRL16.1* was previously mapped on chromosome 16. The *qRL16.1* has been characterized by transcriptome analysis of roots in near-isogenic lines (NILs), gene expression analysis in a pair of lines contrasting with alleles of *qRL16.1*, and differential gene expression analysis in germplasm accessions contrasting with root length. Two candidate genes, *Glyma.16g108500* and *Glyma.16g108700*, have shown relatively higher expression in longer root accessions than in shorter rooting accessions. The C-terminal domain of *Glyma.16g108500* and *Glyma.16g108700* is similar to the conserved domain of C-terminally encoded peptides (CEPs) that regulate root length and nutrient response in Arabidopsis. Two polymorphisms upstream of *Glyma.16g108500* showed a significant association with primary root length and total root length traits in a germplasm set. Synthetic peptide assay with predicted CEP variants of *Glyma.16g108500* and *Glyma.16g108700* demonstrated their positive effect on primary root length. The two genes are root-specific in the early stage of soybean growth and showed differential expression only in the primary root. These genes will be useful for improving soybean to develop a deep and robust root system to withstand low moisture and nutrient regimes.

KEYWORDS

primary root, C-terminally encoded peptides, root length, genes, soybean

1 Introduction

Soybean (*Glycine max* L. Merill) is an important crop grown over an area of 133.8 Mha with a total production of 348.86 Mt during 2022 (https://www.fao.org/faostat/). Soybeans are used in oil, food, feed, nutraceuticals, biofuel, and several non-food industries. Soybean originated in China, and over the period it spread to the American and Asian continents

(Hymowitz, 1970). The developmental biology and genetics of the above-ground parts of soybean have been well characterized, and information on the below-ground parts is slowly emerging. Roots are essential organs of plants for anchoring and fulfilling nutritional requirements, and show remarkable plasticity in response to environmental cues. Root architecture plays an important role in the robust performance of the whole plant by the efficient acquisition of water and nutrients from the soil. To decipher the genetic architecture of root architecture traits in soybean, several quantitative trait loci (QTLs) have been mapped for various root traits (Abdel-Haleem et al., 2011; Brensha et al., 2012; Liang et al., 2014; Manavalan et al., 2015; Prince et al., 2015; Prince et al., 2019; Yang et al., 2019; Prince et al., 2020; Chen et al., 2021; Wang et al., 2022). However, only a few genes have been characterized to control the characteristics of root architecture in soybean. Soybean *Xyloglucan endoglycosylase/hydrolase 1* (*GmXTH1*) overexpression showed higher root phenotypic traits, including main root length, lateral root length, root volume, root dry weight, and root fresh weight in transgenic soybean plants compared to wild-type plants (Song et al., 2022). Similarly, overexpression of two soybean transcription factors, *GmNAC19* and *GmGRAB1*, significantly increased root growth and biomass in transgenic composite plants (Mazarei et al., 2023). Interestingly, the overexpression of *GmNAC19* also enhanced water stress tolerance in transgenic composite plants.

In *Arabidopsis*, genes such as *ROOT GROWTH FACTOR, SHORT-ROOT, SCARECROW, AUXIN1 (AUX1) family*, or efflux carriers of the *PIN-FORMED (PIN) family, EXOCYST70A3, COMPACT ROOT ARCHITECTURE*, and *C-TERMINALLY ENCODED PEPTIDES (CEPs)*, are important genes that regulate root development and nutrient uptake response (Benfey et al., 1993; Scheres et al., 1995; Luschnig et al., 1998; Ruzicka et al., 2007; Swarup et al., 2007; Ohyama et al., 2008; Matsuzaki et al., 2010; Delay et al., 2013; Huault et al., 2014; Ogura et al., 2019). In our previous study, a major locus, *qRL16.1*, was identified to control the length of primary root in soybean (Chen et al., 2021). In our current study, we found that the candidate gene for locus *qRL16.1* have similarity to *Arabidopsis thaliana* C-terminally encoded peptides (CEPs). CEPs are a class of small post-translationally modified signaling peptides that are involved in the regulation of root development (Ohyama et al., 2008; Delay et al., 2013; Roberts et al., 2013). The mature peptide encoded by the *CEP* genes is a 15 amino acid peptide sequence containing two post-translationally modified hydroxyproline residues and is found in the conserved C-terminal domain (Ohyama et al., 2008). CEPs also act as a root-to-shoot mobile peptide hormone that triggers systemic nitrogen acquisition under nitrogen-starved roots in *Arabidopsis* (Ohyama et al., 2008; Tabata et al., 2014; Ohkubo et al., 2017).

The purpose of this study was to characterize the transcriptome of *qRL16.1* and identify best candidate genes through differential expression analysis in soybean germplasm contrasting for primary root length, and association analysis. Synthetic peptide assay and tissue specific gene expression analysis were employed to further support functionality and root specificity of the identified candidate genes. This research will help in improving the soybean to develop a deep root system that can withstand low moisture and nutrient regimes.

2 Materials and methods

2.1 Plant material

K099 (shorter primary root) and Fendou 16 (longer primary root) are parental genotypes of the mapping population used to identify the *qRL16.1* a primary root length locus (Chen et al., 2021). Residual heterozygous line-derived near-isogenic lines (NILs): Near-isogenic lines differing at *qRL16.1*, which were selected from the same recombinant inbred line (RIL) but heterozygous at *qRL16.1* (Chen et al., 2021). NIL-F had the Fendou 16 homozygous genotype, and NIL-K had the K099 homozygous genotype at *qRL16.1*. Advanced backcross lines BC4-F and BC4-K: BC4-F had the Fendou 16 homozygous genotype and BC4-K had the K099 homozygous genotype at the mapped *qRL16.1* region in the background of K099 (Chen et al., 2021). A total of 76 germplasm accessions from a mini-core collection of the world soybean germplasm (Kaga et al., 2012), together with K099 and Fendou16, were used for phenotyping and analysis of the association of root traits with identified polymorphisms.

2.2 RNA-seq of NILs

Root samples of NIL-K and NIL-F were collected from 7-day-old seedlings. Each sample comprised roots collected from five individual plants, and three biological replicates were analyzed. Total RNA was extracted from each sample and cDNA was synthesized with adapters and sequenced using the Illumina HiSeq 2500 analyzer at BGI Technologies (Shenzhen, China). The adapter reads and low-quality reads (bases with quality values ≤ 5) were filtered out of the raw data to obtain high-quality (clean) reads, which were mapped to the soybean reference genome (*G. max* Wm82.a2.v2) downloaded from Phytozome (Goodstein et al., 2012) using HISAT v2.0.4 (http://www.ccb.jhu.edu/software/hisat). Significant differentially expressed genes (DEGs) were obtained with fold changes ≥ 2.0 or ≤ −2.0 for up- or down-regulated genes and adjusted *P*-value ≤ 0.05.

2.3 Sequencing and real-time PCR analysis of the DEGs

Genomic sequences containing coding sequences and upstream promoter regions of two DEGs, *Glyma.16g108500* and *Glyma.16g112400*, were sequenced from both parents, i.e., K099 and Fendou16, using Sanger sequencing. For real-time PCR based expression analysis of DEGs, primary root tip and lateral root tissues from K099, Fendou16, NIL-K, NIL-F, BC4-F, and BC4-K were collected from 3-week-old plants and frozen in liquid nitrogen. RNA isolation from three biological replicates of each sample was performed using the RNeasy plant mini kit (Qiagen, Germany). The isolated RNA was treated with DNAse I using the RapidOut DNA Removal Kit (Thermo Scientific, USA). Purified RNA was used for cDNA synthesis using the PrimeScript RT master mix (TAKARA BIO, Japan). Comparative quantitative real-time PCR (RT-PCR)

10.3389/fpls.2024.1387954

was performed using Powerup SYBR Green Master mix (Applied Biosystems, USA) for selected candidate genes. *GmActin* (*Glyma.18g290800*) was used as an internal reference gene. The sequences of primers used for RT-PCR are given in Table 1. The $2^{-\Delta\Delta CT}$ method was used for analyses of relative gene expression levels (Livak & Schmittgen (2001).

2.4 Phenotyping of world core collection of soybean germplasm

A total of 76 accessions from the world soybean mini-core collection of Japan, along with K099 and Fendou16, were used for phenotyping of primary root length, total root length, root volume, surface area, and root tips using hydroponic culture (Chen et al., 2021). The roots of 15-day-old plants (n = 4) were scanned and analyzed using WinRHIZO software (Regent Instruments Inc., Canada). Phenotyping data from these accessions were used for the marker-trait association analysis. The top four accessions for longer primary roots and the top four accessions for shortest primary roots were used for the validation of four differentially expressed genes.

2.5 Genotyping of SNP and Indel in the germplasm

Accessions of the world soybean mini-core collection of Japan, together with K099 and Fendou16, were genotyped with one SNP (RL1SNP968_T/C) and one Indel (RL1Indel363_–/AA) marker selected from the promoter region of *Glyma.16g108500*. Allele-specific PCR primers were used for the amplification of specific SNP or Indel alleles and were separated on 8.0% polyacrylamide gel. The DNA band pattern was visualized using a Pharos FX™ Molecular Imager (Bio-Rad, Tokyo, Japan) instrument for scoring alleles. The Welch's t-test was used to test the (null) hypothesis that two populations formed by alternate alleles have equal means.

TABLE 1 Primers used for gene expression studies using real-time PCR.

Primer name	Sequence (5′ > 3′)
GmActin-F	ATCTTGACTGAGCGTGGTTATTCC
GmActin-R	GCTGGTCCTGGCTGTCTCC
Glyma.16g108500-F	ACCTTGTGTTGTGCTCTGGT
Glyma.16g108500-R	AGCATGTGGTGGGTGAATGT
Glyma.16g112400-F	ACCACTACACATTGGGTTTGGAT
Glyma.16g112400-R	CTTCCTCACTCTCCCCTGGAT
Glyma.16g108400-F	GGTGATGCTTACCGCCCAA
Glyma.16g108400-R	TGGTGTTTCATGCCCCACA
Glyma.16g108700-F	TGGAAGCATCAAGAAAGCTCA
Glyma.16g108700-R	GGTGTACCAAAGTGCAGGAG

2.6 *CEP*-like genes in soybean

In order to identify *CEP*-like genes in soybean, protein sequences from 16 known *CEP* genes of *Arabidopsis thaliana* were used as queries in a BLASTP search against the Williams 82 genome Wm82.v2 in SoyBase (www.soybase.org). The 14 homologs identified in the soybean genome were used again for iterative BLASTP to identify more *CEP*-like genes in the soybean genome. Thus, 23 more matching gene sequences were identified. The 37 matching sequences were manually filtered for proteins with an N-terminal signal peptide sequence for the secretory pathway and a conserved C-terminal 'CEP-like' sequence motif 'FRPToPGHSPGVGH' (at 70% consensus cutoff), using multiple sequence alignment by CLUSTALW and motif prediction by MEME Suite tool (Bailey et al., 2015). Finally, 25 *CEP*-like genes were selected and used for phylogeny and motif analysis. MEGA11 software was used to construct the phylogeny of 25 *CEP*-like gene protein sequences (Tamura et al., 2021). The neighbor-joining clustering method with 10,000 bootstrap iterations was used to build a consensus phylogenetic tree. The MEME Suite tool was used for consensus motif detection with the following settings: motif site distribution as any number of sites per sequence, motif width 6 to 50, minimum site for motif 2, and maximum number of motifs of 10 (Bailey et al., 2015).

2.7 Synthetic peptide assay

The conserved regions of CEP for *Glyma.16g108500* and its homolog *Glyma.16g108700* were predicted, and putative post-translational modifications of proline amino acid residues, that is, hydroxyproline, were included for synthesis (Table 2). Synthetic peptides *Glyma.16g108500* (CEP-RL1) and *Glyma.16g108700* (CEP-RL2) were synthesized artificially at Eurofins, Japan. The seeds of the short-rooting genotype K099 were sown in vermiculite and young seedlings 4 days after sowing were used for the synthetic peptide assay. The 4-day seedlings were transferred to 0.5× Hoagland solution (pH 6.5). For each treatment, eight seedlings were placed in a plastic box (13 cm × 8.5 cm × 6 cm) containing 500 ml Hoagland solution. The outer surface of the box was covered with aluminum foil to prevent root exposure to the light. Following concentrations were used for the treatment in Hoagland solution: i) Control: No CEP; ii) CEP-RL1: 1 µM; iii) CEP-RL1: 2 µM; iv) CEP-RL2: 1 µM; v) CEP-RL2: 2 µM. The seedlings were incubated in an incubation chamber kept at 22° C night and 24°C day temperatures and supplied with 14 h of fluorescent light. Roots were harvested after 10 days of peptide treatment and scanned for recording of root traits using

TABLE 2 Sequences of synthetic peptides of *Glyma.16g108500* and *Glyma.16g108700* CEPs.

Gene	Peptide name	Peptide sequence
Glyma.16g108500	CEP-RL1	DAFR{HYP}TS{HYP}GHS {HYP}GVGH
Glyma.16g108700	CEP-RL2	DAFR{HYP}TCRGHS{HYP}GAGH

WinRHIZO (Regent Instruments Inc., Canada). The recorded root traits were primary root length, total root length, and number of root tips.

2.8 Tissue-specific gene expression of *Glyma.16g108500* and *Glyma.16g108700*

Two time stages, 4-day, and 7-day, were selected for gene expression analysis of *Glyma.16g108500* and *Glyma.16g108700* in parental genotypes K099 and Fendou16. Tissues were collected at 4 days after sowing, from the shoot apex (2 mm) and primary root tip (2 cm), and at 7 days after sowing, from shoot apex (2 mm), the primary root tip (2 cm), and lateral root tips (2 cm), for tissue-specific gene expression analysis.

3 Results

3.1 Identification of DEGs for *qRL16.1* by root transcriptome analysis

RNA-Seq analysis of the root transcriptome of two NILs identified total 317 genes upregulated and 192 downregulated in NIL-F compared to NIL-K (Supplementary Figure 1A). Differential expression analysis for genes from the genomic region of primary root length QTL *qRL16.1* (Chr.16:27158682.28953489), among two NILs, identified only two genes, *Glyma.16g108500* and *Glyma.16g112400*, showing a significant fold difference in their expression levels (Supplementary Table 1). The gene expression analysis by RT-PCR identified that the expression of *Glyma.16g108500* was 3.0 fold higher in Fendou16 tissues compared to K099 tissues, and it was 4.47 fold higher in NIL-F tissues than in NIL-K tissues. Conversely, the expression level of *Glyma.16g112400* was 18.11 fold lower in Fendou16 tissues compared to K099 tissues, and it was 7.21 fold lower in NIL-F tissues than in NIL-K tissues (Supplementary Figures 1B, C). The genomic sequence of two DEGs, *Glyma.16g108500* and *Glyma.16g112400*, including a 2-kb promoter region, was sequenced from both parents, Fendou16 and K099. No sequence differences were identified in the coding DNA sequence of both genes between the two parents. Upstream 2-kb sequencing of *Glyma.16g108500* identified two base insertions at the 363 position (–/TA), a single nucleotide polymorphism (SNP) at -968 (T/C), and a single base insertion at the 1038 position (A) in Fendou 16. The 2 kb upstream sequencing of *Glyma.16g112400* did not detect any sequence difference between the two parental genotypes.

3.2 Gene annotations and homology search identified *Glyma.16g108500* as a *CEP*-like gene in soybean

According to gene annotations assigned in SoyBase (www.soybase.org), *Glyma.16g108500* was annotated as an uncharacterized protein, root development, and DNA-directed RNA polymerase II subunit RPB1-like protein, while *Glyma.16g112400* was annotated as an uncharacterized protein. Furthermore, the analysis of protein homology in the InterPro protein database identified that *Glyma.16g108500* encodes a C-terminally encoded peptide (CEP), a C-terminally conserved secreted protein in *Arabidopsis* (https://www.ebi.ac.uk/interpro/entry/InterPro/IPR033250/). Based on protein homology, two more paralogs of *Glyma.16g108500* were also identified in close proximity to *Glyma.16g108500* in the genomic region of *qRL16.1*. These two paralogs, *Glyma.16g108700* and *Glyma.16g108400*, showed 72% and 80% homology of the protein sequence with *Glyma.16g108500*, respectively.

3.3 Validation of identified DEGs in parents, NILs, and advanced backcross lines

The two genes *Glyma.16g108500* and *Glyma.16g112400* were validated for differential expression by real-time PCR in primary root tip of two parents, a pair of contrasting NILs, and a pair of contrasting BC_4 backcross lines. Expression analysis identified that *Glyma.16g108500* showed higher expression in longer rooting genotypes, whereas *Glyma.16g112400* showed higher expression in shorter rooting genotypes (Figure 1). In the case of *Glyma.16g108500*, although the expression fold difference in the contrasting alleles NILs and BC_4s was similar to that of the parents, the level of expression in both pairs of contrasting lines was slightly reduced compared with that in the two parents. The analysis of gene expression for two *Glyma.16g108500* paralogs, *Glyma.16g108700* and *Glyma.16g108400*, among three pairs of contrasting genotypes identified that *Glyma.16g108700* and *Glyma.16g108400* also showed higher expression in longer rooting genotypes; however, the level of expression fold difference was relatively higher in the case of *Glyma.16g108700* (Figure 1).

3.4 Phenotyping of the world soybean mini-core collection for root traits

The descriptive statistics of different root traits in 76 soybean accessions, Fendou 16 and K099, at 15 days of growth are given in Table 3. The primary root length ranged from 16.3 cm in CHUUHOKU 2 to 51.45 cm in BARITOU 3A, with a mean root length of 34.4 cm. The total root length varied from 142.8 cm to 606.4 cm with a mean value of 402.7 cm.

The five high and five low contrasting accessions for primary root length and total root length traits along with Fendou 16 and K099 were again evaluated. In this confirmation experiment, similar results were obtained for short- and long-rooting genotypes. Among the top five accessions for primary root length, BARITOU 3A showed the highest primary root length of 53.1 cm, followed by RINGGIT (51 cm) and Fendou 16 (49.2 cm). From this experiment, the top four accessions for longer primary root and the four best accessions for short primary root were selected for validation of four differentially expressed genes (Table 4, Figure 2).

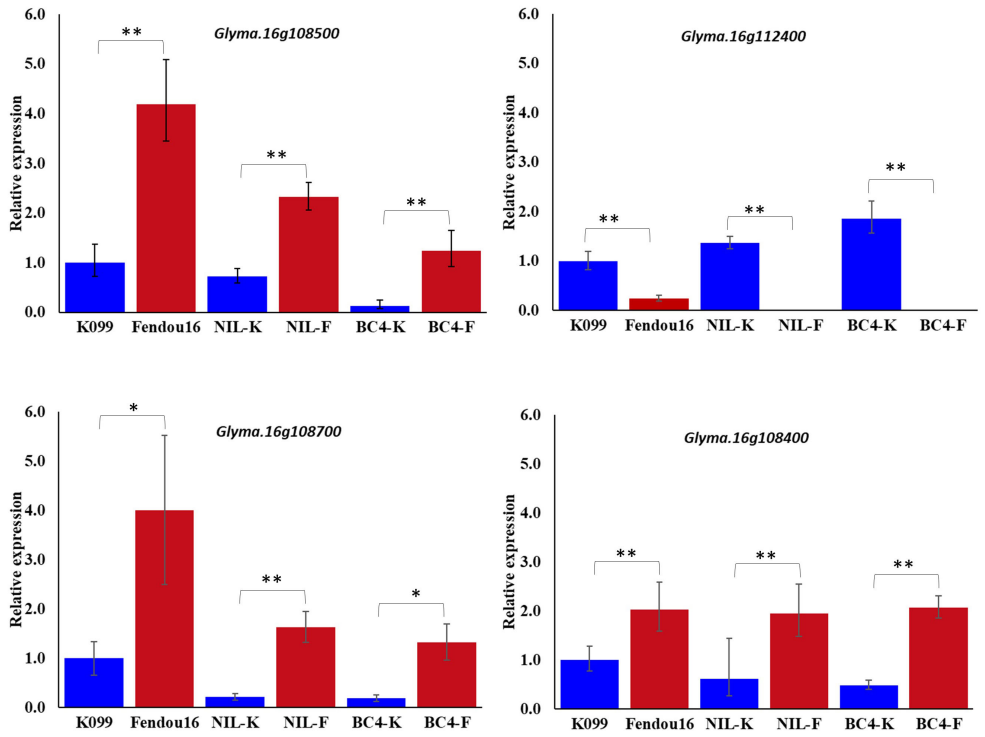


FIGURE 1
Relative expression of *Glyma.16g108500*, *Glyma.16g112400*, *Glyma.16g108700* and *Glyma.16g108400*, at 3 weeks primary root tips of K099, Fendou 16, NILs and BC_4 backcross lines. *GmActin* was used as a normalizer. Significant differences between K099 and Fendou 16, NIL-K and NIL-F, BC4-K and BC4-F, are marked with asterisks (*P<0.05; **P<0.01; Student's t-test).

3.5 Validation of identified DEGs in contrasting root-length germplasm

The four differentially expressed genes identified among three contrasting pairs of high and low root length lines from the genomic region of *qRL16.1* were further analyzed for their expression in the top four long rooting and top-four shortest rooting germplasm selected from the world core collection of soybean. Gene expression analysis using real-time PCR showed that *Glyma.16g108500* and *Glyma.16g108700* had relatively higher expression in longer rooting genotypes than in shorter rooting accessions, except for accession Heukdaelip (Figure 3). However, *Glyma.16g108400* and *Glyma.16g112400* did not show differential expression between the two contrasting groups in accordance with root length and expression pattern observed in the NILs, thus demeaning their role in the regulation of the development of primary root in soybean (Figure 3).

3.6 Association analysis in world soybean mini-core collection germplasm

Coding region of both DEGs, *Glyma.16g108500* and *Glyma.16g108700*, was sequenced but no polymorphism was detected between K099 and Fendou 16. The two polymorphisms, SNP at -968 (RL1SNP968_T/C) and Indel at -363 (RL1INDEL363_–/TA), identified in the upstream sequence of *Glyma.16g108500*, were analyzed in the germplasm to test the association of allelic variation present in the proximity of two DEGs identified in contrasting

TABLE 3 Descriptive statistics of root traits in 78 soybean accessions of the soybean mini core collection (Japan).

	Primary root length (cm)	Total root length (cm)	Total root volume (cm^3)	Total root tips	Root surface area (cm^2)
Minimum	16.30	142.87	0.41	60.75	27.15
Maximum	51.45	606.48	1.74	450.50	108.53
Mean	34.40	402.76	0.98	257.46	69.86
Standard deviation	6.41	111.87	0.29	82.79	18.85
Variance	41.09	12515.11	0.09	6854.27	355.19

10.3389/fpls.2024.1387954

TABLE 4 Primary root length and total root length values of top four contrasting accessions along with K099 and Fendou 16.

	Genotypes	Primary root length (cm)	Total root length (cm)
Top five longest rooting type	BARITOU 3A	53.12	475.48
	RINGGIT	51.05	552.40
	Fendou 16	49.25	468.78
	MERAPI	47.78	636.60
	HM39	46.20	489.32
Top four shortest rooting genotypes along with K099	KE32	23.43	229.18
	U 1155-4	25.38	422.28
	CHUUHOKU 2	25.50	367.38
	HEUKDAELIP	28.85	390.67
	K099	37.37	523.68

germplasm. Genotyping of the RL1SNP968_T/C and RL1INDEL363_–/TA markers showed identical marker genotypes among 78 genotypes with 36% and 64% allele frequencies for the K099 and Fendou16 alleles, respectively. The Welch *t*-test showed significant differences in primary root length and total root length between two populations formed by alternate alleles of RL1INDEL363_–/TA (Table 5, Figure 4).

3.7 *CEP*-like genes in soybean

A total of 25 *CEP*-like genes were identified that contain an N-terminal signal peptide and the CEP-like conserved motif 'FRPToPGHSPGVGH' (at consensus cutoff of 70%). These 25 *CEP*-like genes were distributed on seven chromosomes. Except for Chr.06 and 09, other *CEP*-like genes were located in tandem, suggesting that their copy number increased by tandem duplication. The highest tandem copies (eight) were present on Chr. 17. Phylogeny analysis using the neighbor-joining clustering method showed that most CEP-like proteins were clustered together in group I, while two CEP-like proteins from Chr.01 and Chr.09 were grouped in cluster II. The two CEP-like proteins of *Glyma.16g108400* and *Glyma.16g108500* from Chr.16 were clustered in group III, whereas the CEP protein from *Glyma.16g108700* was most divergent (Figure 5). Consensus motif analysis using MEME Suite identified consensus motif distribution patterns among 25 CEP proteins with an N-terminal signal peptide motif having only one site for each sequence, while a C-terminally encoded peptide conserved motif (Motif 1) was present as a single copy in 19 members and as two copies in six members (Figure 6). Two conserved domains (VLWYTPND and MINIHPPPAIPRSPQ) are specifically found in three *CEP*-like genes of Chr.16 and two members of group II of the phylogenetic tree, that is, *Glyma.01G152800* and *Glyma.09G218000* (Figure 6). Three proline residues were present in the 15 amino acid CEP domain of *Glyma.16g108500*, while only two proline residues were present in the CEP domain of *Glyma.16g108700*.

3.8 Synthetic peptide experiments

Because the analysis of gene expression in contrasting germplasm excluded *Glyma.16g112400* and *Glyma.16g108400* from the possibility of their positive role in regulating the development of primary root length, we focused on *Glyma.16g108500* and *Glyma.16g108700* for further studies using 15 amino acid synthetic peptide of the predicted CEP domain for these genes. Two different

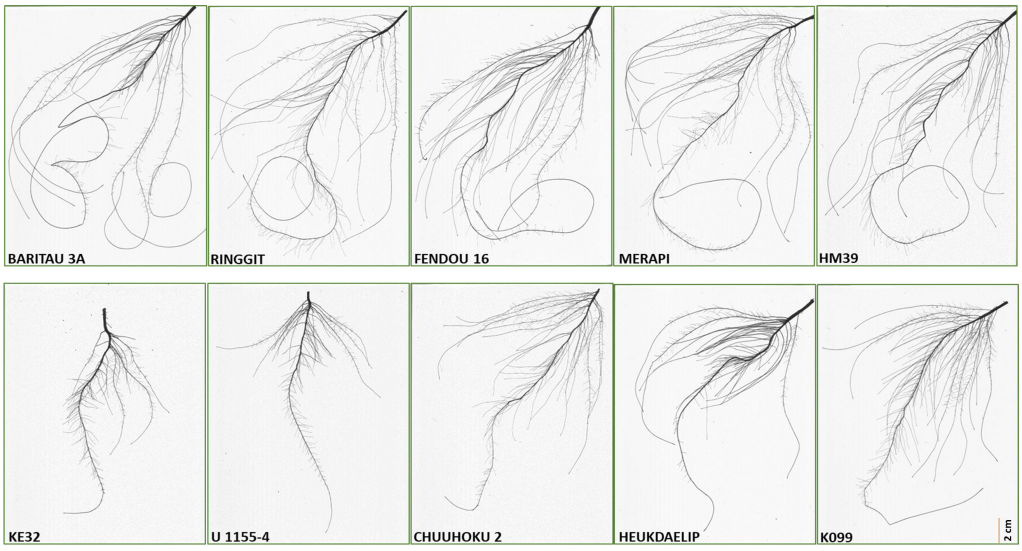


FIGURE 2
Root scan pictures of the top four contrasting root length genotypes along with K099 and Fendou 16.

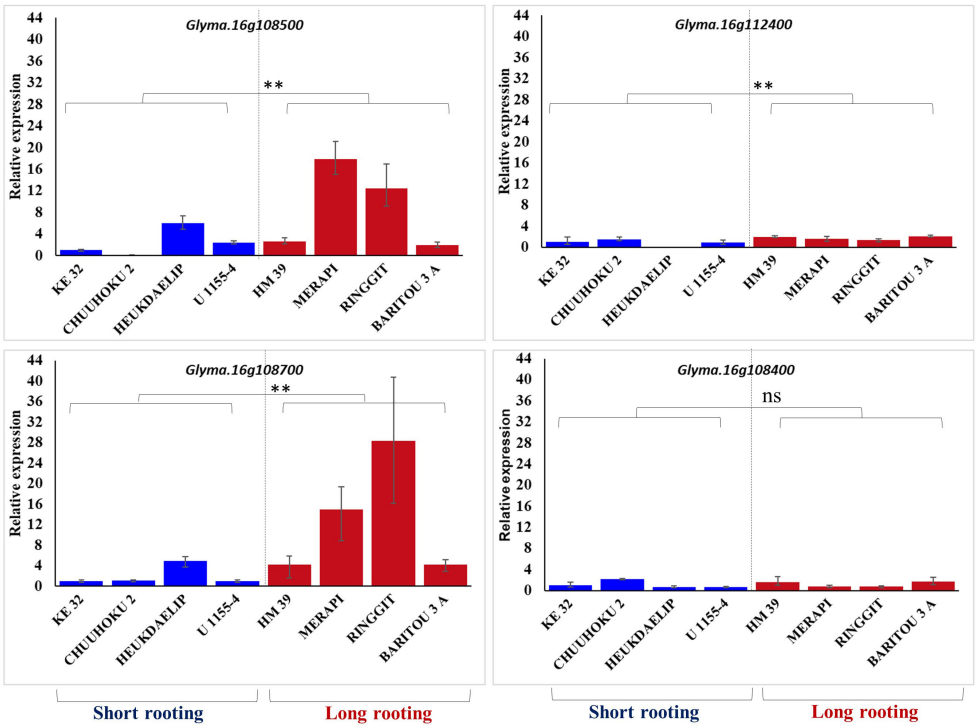


FIGURE 3
Relative expression of *Glyma.16g108500*, *Glyma.16g112400*, *Glyma.16g108700*, and *Glyma.16g108400* in the four genotypes with the longest and shortest rooting genotypes. *GmActin* was used as a normalizer. Significant differences between short rooting and long rooting genotype sets are marked with asterisks (**P<0.01; ns, not significant; Student's t-test).

concentrations of two CEPs for these genes (CEP-RL1 and CEP-RL2) were applied in Hoagland solution-based hydroponic culture (pH 6.5) for the short-rooting genotype K099. The results of this experiment showed that the length of the primary root was significantly higher for CEP-RL1 at concentrations of 1 μM and 2 μM s, whereas for CEP-RL2, the length of the primary root was significantly higher at a concentration of 2 μM concentration, compared to the control (Figure 7). Analysis of the effect of CEP treatment on total root length and root tips showed that at higher concentrations of each CEP (2 μM), total root length and root tips were significantly lower in CEP-RL1 treatment, but not in the case of CEP-RL2 treatment (Supplementary Figure 2).

TABLE 5 Welch's t-test for RL1INDEL363_−/TA allele association with root length traits in 78 germplasm accessions.

RL1INDEL363_−/TA allele	Primary root length (cm)		Total root length (cm)	
	Allele A	Allele B	Allele A	Allele B
Mean	32.62	36.02	349.88	442.33
Variance	60.99	26.47	12218	8225
Observations (*N*)	28	50	28	50
t-Statistics	-2.066		-3.771	
P(T<=t) one-tail	0.02267		0.00022	
t Critical one-tail	1.6838		1.6779	
P(T<=t) two-tail	0.04534*		0.00045 **	
t Critical two-tail	2.02107		2.01174	

[#]Allele A represent K099 allele and Allele B represent Fendou 16 allele. * and ** shows significant at P < 0.05 and P < 0.001 in Welch's t-test, respectively.

3.9 Tissue-specific gene expression analysis

In order to observe the expression pattern of two CEP genes in meristematic tissues of parental genotypes, the shoot apex and primary root tip were analyzed 4-day after sowing. Lateral roots appeared properly after 6–7 days of sowing, and so the shoot tip, the primary root tip, and the lateral root tips were analyzed at the 7-day stage. The analysis of gene expression for *Glyma.16g108500* in the shoot and the primary root tip of 4-day seedlings showed expression only in the primary root tip, where the expression was 2.7-fold higher in Fendou 16 than in K099 (Figure 8A). For *Glyma.16g108700*, expression was observed in the primary root, but the expression level was similar in both genotypes (Figure 8A). No expression was observed in shoot apex for either gene. These results indicated that the genes *Glyma.16g108500* and *Glyma.16g108700* are root-specific in the early stage of root development. In the primary root and lateral root tissues of 7-day seedlings, both genes showed expression and the expression level was higher in the primary root than in the lateral roots (Figure 8B).

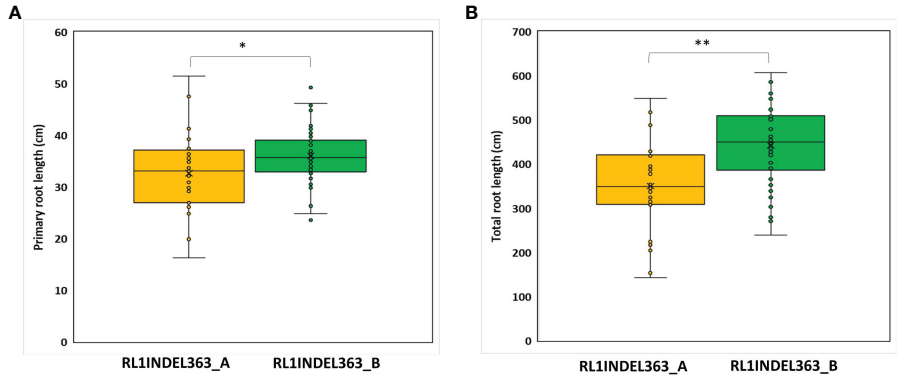


FIGURE 4
Genotypic association for RL1INDEL363_−/TA alleles with root traits in 78 accessions of the mini core collection of Japan **(A)** Primary root length, **(B)** Total root length. RL1INDEL363_A represents the K099 allele and RL1INDEL363_B represents the Fendou 16 allele. * and ** show significant at $P < 0.05$ and $P < 0.001$ in the Welch's t-test, respectively.

Fendou 16 showed a 3.2-fold higher expression of *Glyma.16g108500* and 1.8-fold higher expression of *Glyma.16g108700* in primary root than K099, but no differential expression was observed in lateral roots. In the shoot apex of 7-day seedlings, no expression was observed for *Glyma.16g108500*, but low expression was observed for *Glyma.16g108700*. These results also showed the involvement of these two genes in the early stage of primary root development.

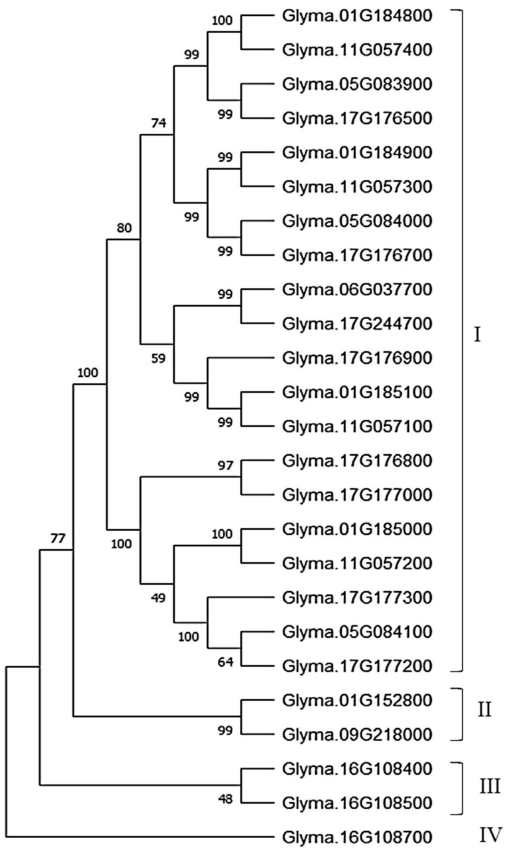


FIGURE 5
Phylogenetic tree of the protein sequence of 25 *CEP*-like genes in *Glycine max*..

4 Discussion

Genes controlling root traits are difficult to characterize due to the difficulty in phenotyping root traits under field conditions. Furthermore, roots exhibit high plasticity even under slight changes in the growing environment, making accurate phenotyping more difficult. Genetic architecture of root architecture traits in soybean has been dissected in some studies by root phenotyping under hydroponic, semi-hydroponic and field conditions, and quantitative trait loci (QTLs) mapping (Abdel-Haleem et al., 2011; Brensha et al., 2012; Liang et al., 2014; Manavalan et al., 2015; Prince et al., 2015; Yang et al., 2019; Prince et al., 2020; Chen et al., 2021; Wang et al., 2022). In our previous study a major locus for primary root length, *qRL16.1*, was mapped on Chr.16 explaining 30.25% of the total phenotypic variation under hydroponic culture (Chen et al., 2021). Traditionally, genes underlying QTLs were cloned and characterized using map-based cloning, which is a time-consuming process. However, in the next-generation sequencing (NGS) era and post NGS-era, gene identification and characterization are much faster (Jaganathan et al., 2020). We used an approach similar to bulk segregant RNA Seq (BSR-Seq) for gene discovery (Liu et al., 2012). In our current study, complete transcriptome sequencing of contrasting NILs along with their parents was performed to identify DEGs in the *qRL16.1* genomic region. The DEGs identified in this way were further analyzed for their association with the primary root length trait in contrasting germplasm using RT-PCR and marker-trait association analysis. The germplasm accessions of the high rooting group in the contrasting germplasm set had an almost two-fold difference in primary root length compared to the low rooting group. The contrasting NILs and advanced backcross lines pairs showed differential expression for all four genes tested; however, the contrasting germplasm accessions set showed differential expression only for two genes, *Glyma.16g108500* and *Glyma.16g108700;* this narrowed down further work to these two

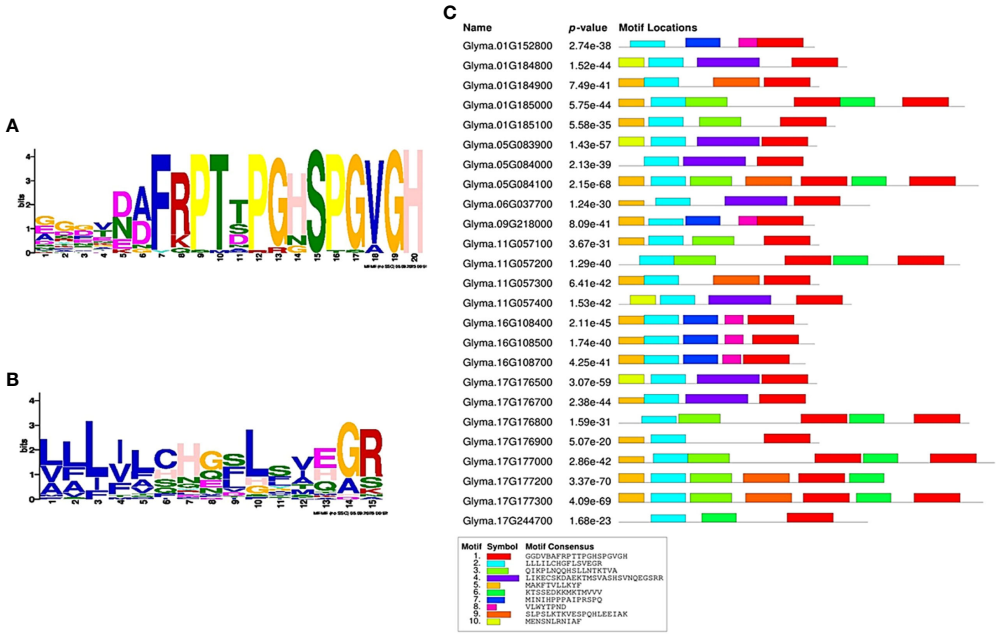


FIGURE 6
Consensus motif pattern among 25 CEP-like proteins in *Glycine max*. **(A)** Weblogo of the CEP consensus motif (Motif 1), **(B)** Weblogo of the N-terminal signal peptide consensus motif (Motif 2), **(C)** Consensus motif composition of 25 CEP-like proteins detected by MEME Suite.

genes only. The SNP and Indel polymorphisms, RL1SNP968_T/C and RL1INDEL363_–/TA, present upstream of *Glyma.16g108500* had identical genotypes in all germplasm analyzed and constitute a haplotype. The frequency of the Fendou16 allele was higher in this haplotype and it showed a significant association with the length of the primary root and total roots. Finally, the function of these genes was validated by external treatment of synthetic peptides of the CEP domains of the two secretary peptide genes, which demonstrated their effect on primary root length.

Secreted regulatory peptides are signaling molecules that arise from genes that typically encode an N-terminal secretion signal and one or more conserved peptide domains (Delay et al., 2013). They are involved in many aspects of the development of shoots and roots. The conserved C-terminal domains of *Glyma.16g108500* and *Glyma.16g108700* are similar to the conserved domains of Group-I *CEP*-like genes in *Arabidopsis* (Roberts et al., 2013). The *CEP* gene family is known for its role in root length development and nutrient response in Arabidopsis (Ohyama et al., 2008; Delay et al., 2013; Roberts et al., 2013). The two *CEP*-like genes identified in this study for root development in soybean have a conserved C-terminal domain and an N-terminal signal peptide for the secretory pathway. Interestingly, in the case of *Arabidopsis CEP*-like genes,

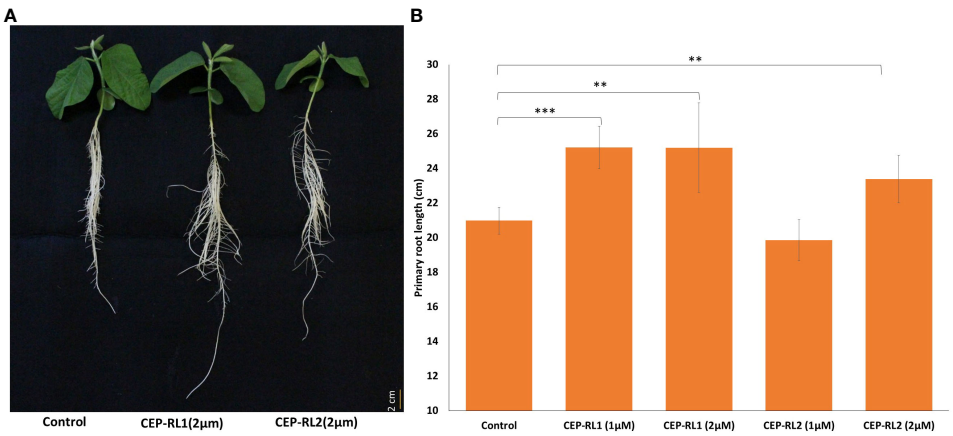


FIGURE 7
Effect of synthetic C-terminally encoded peptides (CEPs) on the development of the primary root length in K099, **(A)**. Shoot and root phenotypes of control and CEP treated plants (0. 5× Hogland solution pH 6.5), **(B)**. The difference in primary root length between control and CEP-treated plants (n = 8). *** and ** indicate significant differences at $P < 0.001$, and $P < 0.01$, respectively, as identified by the Student's t-test.

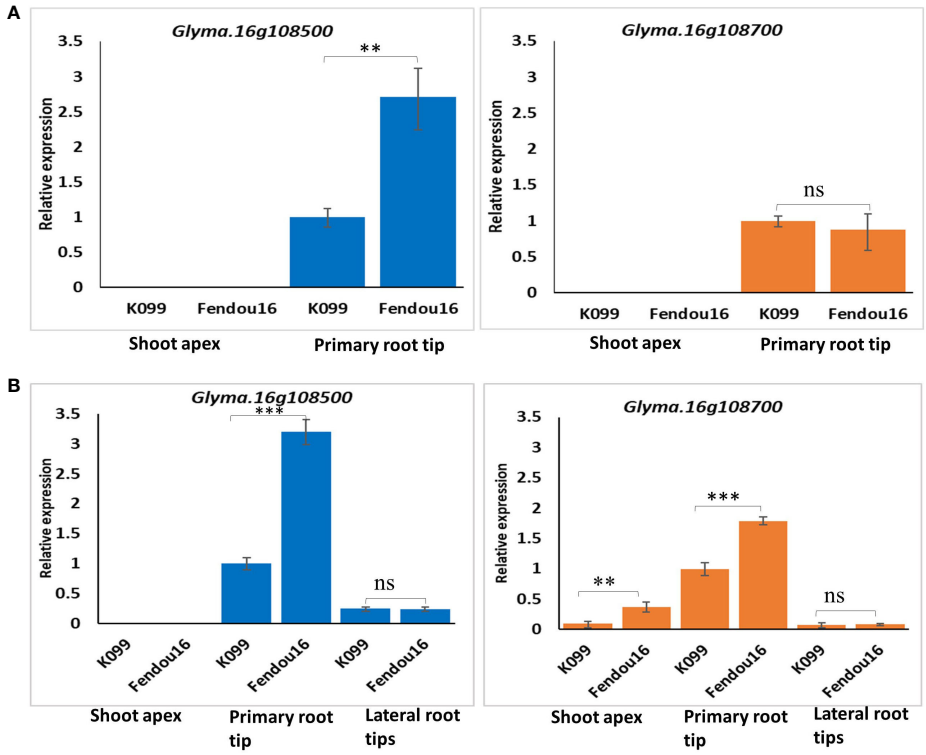


FIGURE 8
Relative expression of *Glyma.16g108500* and *Glyma.16g108700* in K099 and Fendou 16, **(A)** 4-day tissues, **(B)** 7-day tissues. *GmActin* was used as a normalizer. Significant differences between K099 and Fendou 16, are marked with asterisks (**P<0.01; ***P<001; ns, not significant; Student's t-test).

overexpression or external application of CEPs has shown reduced root growth (Ohyama et al., 2008; Roberts et al., 2013). In Arabidopsis, CEPs reduce the length of the primary root by slowing growth and reducing the lateral root density prior to the initiation of the lateral root (Delay et al., 2013).

Studies on the negative effect of increasing the concentration of CEP on root development have been reported in *Arabidopsis* (Ohyama et al., 2008; Delay et al., 2013; Roberts et al., 2013), Apple (Yu et al., 2019) and *Medicago* (Imin et al., 2013); however, our study is the first report on the positive effect of a higher concentration of CEPs on root length. During evolution, the soybean genome has undergone two rounds of genome duplication events, and subsequent gene duplication and neofunctionalization resulted in the acquisition of new functions for many genes (Schmutz et al., 2010). This may be the reason why higher expression of *CEP*-like genes is a positive regulator of root length in soybean. The number of *CEP* genes in soybean is also higher than that reported in any other species (Delay et al., 2013; Imin et al., 2013; Roberts et al., 2013; Yu et al., 2019; Aggarwal et al., 2020). There may be a distinct and important role for different CEPs in soybean. Further experiments on CEP overexpression, gene knockout using gene editing, and external application of different *CEP*-like proteins in soybean may reveal their functionality.

In conclusion, transcriptome analysis, gene expression studies in contrasting NILs and germplasm, marker-trait association in germplasm, and synthetic peptide assay provided strong evidence for the involvement of the two *CEP*-like genes *Glyma.16g108500* and *Glyma.16g108700* in the regulation of root length in soybean. These genes will be useful for soybean improvement toward developing a deep and robust root system to withstand low moisture and nutrient regimes.

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

GK: Writing – original draft, Investigation, Formal analysis. DC: Writing – review & editing, Investigation, Formal analysis. CP: Writing – review & editing, Investigation. DX: Writing – review & editing, Supervision, Funding acquisition, Conceptualization.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported in-part by the Japan International Research Center for Agricultural Sciences under a research project "Resilient Crops" and a JSPS KAKENHI Grant Number 23K05172.

Acknowledgments

We thank Ms. Nozawa, Ms. Kawashima, and Ms. Xiao for their assistances in the experiments.

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.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2024.1387954/full#supplementary-material

References

Abdel-Haleem, H., Lee, G. J., and Boerma, R. H. (2011). Identification of QTL for increased fibrous roots in soybean. *Theor. Appl. Genet.* 122, 935–946. doi: 10.1007/s00122-010-1500-9

Aggarwal, S., Kumar, A., Jain, M., Sudan, J., Singh, K., Kumari, S., et al. (2020). C-terminally encoded peptides (CEPs) are potential mediators of abiotic stress response in plants. *Physiol. Mol. Biol. Plants* 26, 2019–2033. doi: 10.1007/s12298-020-00881-4

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

Benfey, P. N., Linstead, P. J., Roberts, K., Schiefelbein, J. W., Hauser, M. T., and Aeschbacher, R. A. (1993). Root development in Arabidopsis: four mutants with dramatically altered root morphogenesis. *Development* 119, 57–70. doi: 10.1242/dev.119.1.57

Brensha, W., Kantartzi, S. K., Meksem, K., Grier, R. L., Barakat, A., Lightfoot, D. A., et al. (2012). Genetic analysis of root and shoot traits in the 'Essex' by 'Forrest' recombinant inbred line population of soybean. *J. Plant Genom. Sci.* 1, 1–9. doi: 10.5147/jpgs.2012.0051

Chen, H., Kumawat, G., Yan, Y., Fan, B., and Xu, D. (2021). Mapping and validation of a major QTL for primary root length of soybean seedlings grown in hydroponic conditions. *BMC Genomics* 22, 132. doi: 10.1186/s12864-021-07445-0

Delay, C., Imin, N., and Djordjevic, M. A. (2013). *CEP* genes regulate root and shoot development in response to environmental cues and are specific to seed plants. *J. Exp. Bot.* 64, 5383–5394. doi: 10.1093/jxb/ert332

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

Huault, E., Laffont, C., Wen, J., Mysore, K. S., Ratet, P., Duc, G., et al. (2014). Local and systemic regulation of plant root system architecture and symbiotic nodulation by a receptor-like kinase. *PloS Genet.* 10, e1004891. doi: 10.1371/journal.pgen.1004891

Hymowitz, T. (1970). On the domestication of the soybean. *Econ. Bot.* 24, 408–421. doi: 10.1007/BF02860745

Imin, N., Mohd-Radzman, N. A., Ogilvie, H. A., and Djordjevic, M. A. (2013). The peptide-encoding *CEP1* gene modulates lateral root and nodule numbers in *Medicago truncatula*. *J. Exp. Bot.* 64, 5395–5409. doi: 10.1093/jxb/ert369

Jaganathan, D., Bohra, A., Thudi, M., and Varshney, R. K. (2020). Fine mapping and gene cloning in the post-NGS era: advances and prospects. *Theor. Appl. Genet.* 133, 1791–1810. doi: 10.1007/s00122-020-03560-w

Kaga, A., Shimizu, T., Watanabe, S., Tsubokura, Y., Katayose, Y., Harada, K., et al. (2012). Evaluation of soybean germplasm conserved in NIAS genebank and development of mini core collections. *Breed. Sci.* 61, 566–592. doi: 10.1270/jsbbs.61.566

Liang, H., Yu, Y., Yang, H., Xu, L., Dong, W., Du, H., et al. (2014). Inheritance and QTL mapping of related root traits in soybean at the seedling stage. *Theor. Appl. Genet.* 127, 2127–2137. doi: 10.1007/s00122-014-2366-z

Liu, S., Yeh, C. T., Tang, H. M., Nettleton, D., and Schnable, P. S. (2012). Gene mapping via bulked segregant RNA-Seq (BSR-Seq). *PloS One* 7, e36406. doi: 10.1371/journal.pone.0036406

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

Luschnig, C., Gaxiola, R. A., Grisafi, P., and Fink, G. R. (1998). *EIR1*, a root specific protein involved in auxin transport, is required for gravitropism in *Arabidopsis thaliana*. *Genes Dev.* 12, 2175–2187. doi: 10.1101/gad.12.14.2175

Manavalan, L. P., Prince, S. J., Musket, T. A., Chaky, J., Deshmukh, R., Vuong, T. D., et al. (2015). Identification of novel QTL governing root architectural traits in an interspecific soybean population. *PloS One* 10, e0120490. doi: 10.1371/journal.pone.0120490

Matsuzaki, Y., Ogawa-Ohnishi, M., Mori, A., and Matsubayashi, Y. (2010). Secreted peptide signals required for maintenance of root stem cell niche in *Arabidopsis*. *Science* 329, 1065–1067. doi: 10.1126/science.1191132

Mazarei, M., Routray, P., Piya, S., Stewart, C. N. Jr., and Hewezi, T. (2023). Overexpression of soybean *GmNAC19* and *GmGRAB1* enhances root growth and water-deficit stress tolerance in soybean. *Front. Plant Sci.* 14. doi: 10.3389/fpls.2023.1186292

Ogura, T., Goeschl, C., Filiault, D., Mirea, M., Slovak, R., Wolhrab, B., et al. (2019). Root system depth in Arabidopsis is shaped by *EXOCYST70A3* via the dynamic modulation of auxin transport. *Cell* 178, 400–412. doi: 10.1016/j.cell.2019.06.021

Ohkubo, Y., Tanaka, M., Tabata, R., Ogawa-Ohnishi, M., and Matsubayashi, Y. (2017). Shoot-to-root mobile polypeptides involved in systemic regulation of nitrogen acquisition. *Nat. Plants* 3, 17029. doi: 10.1038/nplants.2017.29

Ohyama, K., Ogawa, M., and Matsubayashi, Y. (2008). Identification of a biologically active, small, secreted peptide in Arabidopsis by in *silico* gene screening, followed by LC-MS-based structure analysis. *Plant J.* 55, 152–160. doi: 10.1111/j.1365-313X.2008.03464.x

Prince, S. J., Song, L., Qiu, D., Maldonado Dos Santos, J. V., Chai, C., Joshi, T., et al. (2015). Genetic variants in root architecture-related genes in a *Glycine soja* accession, a potential resource to improve cultivated soybean. *BMC Genomics* 16, 132. doi: 10.1186/s12864-015-1334-6

Prince, S. J., Valliyodan, B., Ye, H., Yang, M., Tai, S., Hu, W., et al. (2019). Understanding genetic control of root system architecture in soybean: Insights into the genetic basis of lateral root number. *Plant Cell Environ.* 42, 212–229. doi: 10.1111/pce.13333

Prince, S. J., Vuong, T. D., Wu, X., Bai, Y., Lu, F., Kumpatla, S. P., et al. (2020). Mapping quantitative trait loci for soybean seedling shoot and root architecture traits in an inter-specific genetic population. *Front. Plant Sci.* 11, 1284. doi: 10.3389/fpls.2020.01284

Roberts, I., Smith, S., De Rybel, B., Van Den Broeke, J., and Smet, W. (2013). The CEP family in land plants: evolutionary analyses, expression studies, and role in *Arabidopsis* shoot development. *J. Exp. Bot.* 64, 5371–5381. doi: 10.1093/jxb/ert331

Ruzicka, K., Ljung, K., Vanneste, S., Podhorska, R., Beeckman, T., Friml, J., et al. (2007). Ethylene regulates root growth through effects on auxin biosynthesis and transport-dependent auxin distribution. *Plant Cell* 19, 2197–2212. doi: 10.1105/tpc.107.052126

Scheres, B., Di Laurenzio, L., Willemsen, V., Hauser, M.-T., Janmaat, K., Weisbeek, P., et al. (1995). Mutations affecting the radial organisation of the Arabidopsis root display specific defects throughout the radial axis. *Development* 121, 53–62. doi: 10.1242/dev.121.1.53

Schmutz, J., Cannon, S., Schlueter, J., Ma, J., Mitros, T., Nelson, L., et al. (2010). Genome sequence of the palaeopolyploid soybean. *Nature* 463, 178–183. doi: 10.1038/nature08670

Song, Y., Zhang, Y., Ye-yao Sujie Fan, D., Di Qin, D., Zhang, Z., and Wang, P. (2022). The role of *GmXTH1*, a new xyloglucan endotransglycosylase/hydrolase from soybean, in regulating soybean root growth at seedling stage. *bioRxiv*. doi: 10.1101/2022.07.15.500193. 2022.07.15.500193.

Swarup, R., Perry, P., Hagenbeek, D., van der Straeten, D., Beemster, G. T., Sandberg, G., et al. (2007). Ethylene upregulates auxin biosynthesis in Arabidopsis seedlings to enhance inhibition of root cell elongation. *Plant Cell* 19, 2186–2196. doi: 10.1105/tpc.107.052100

Tabata, R., Sumida, K., Yoshii, T., Ohyama, K., Shinohara, H., and Matsubayashi, Y. (2014). Perception of root-derived peptides by shoot LRR-RKs mediates systemic N-demand signaling. *Science* 346, 343–346. doi: 10.1126/science.1257800

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

Wang, Z., Huang, C., Niu, Y., Yung, W. S., Xiao, Z., Wong, F. L., et al. (2022). QTL analyses of soybean root system architecture revealed genetic relationships with shoot-related traits. *Theor. Appl. Genet.* 135, 4507–4522. doi: 10.1007/s00122-022-04235-4

Yang, L., Lv, H., and Liao, H. (2019). Identification and mapping of two independent recessive loci for the root hairless mutant phenotype in soybean. *Theor. Appl. Genet.* 132, 301–312. doi: 10.1007/s00122-018-3217-0

Yu, Z., Xu, Y., Liu, L., Guo, Y., Yuan, X., Man, X., et al. (2019). The importance of conserved serine for C-terminally encoded peptides function exertion in apple. *Int. J. Mol. Sci.* 20, 775. doi: 10.3390/ijms20030775

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 01 July 2024
DOI 10.3389/fpls.2024.1403841

OPEN ACCESS

EDITED BY
Yi-Hong Wang,
University of Louisiana at Lafayette,
United States

REVIEWED BY
Muhammad Awais Farooq,
University of Bologna, Italy
Vishnu Mishra,
University of Delaware, United States
Kazuo Nakashima,
Japan International Research Center for Agricultural Sciences (JIRCAS), Japan

*CORRESPONDENCE
Li Song
songli@yzu.edu.cn

†These authors have contributed equally to this work

RECEIVED 20 March 2024
ACCEPTED 12 June 2024
PUBLISHED 01 July 2024

CITATION
Geng X, Dong L, Zhu T, Yang C, Zhang J, Guo B, Chen H, Zhang Q and Song L (2024) Genome-wide analysis of soybean hypoxia inducible gene domain containing genes: a functional investigation of GmHIGD3. *Front. Plant Sci.* 15:1403841. doi: 10.3389/fpls.2024.1403841

Genome-wide analysis of soybean hypoxia inducible gene domain containing genes: a functional investigation of GmHIGD3

Xiaoyan Geng[1†], Lu Dong[1†], Tiantian Zhu[1], Chunhong Yang[1], Jianhua Zhang[1], Binhui Guo[1], Huatao Chen[2,3], Qun Zhang[4] and Li Song[1,2*]

[1]Joint International Research Laboratory of Agriculture and Agri-Product Safety, the Ministry of Education of China, Jiangsu Key Laboratory of Crop Genomics and Molecular Breeding, Yangzhou University, Yangzhou, Jiangsu, China, [2]Zhongshan Biological Breeding Laboratory, Nanjing, China, [3]Institute of Industrial Crops, Jiangsu Academy of Agricultural Sciences, Nanjing, China, [4]College of Life Sciences, State Key Laboratory of Crop Genetics & Germplasm Enhancement and Utilization, Nanjing Agricultural University, Nanjing, China

The response of Hypoxia Inducible Gene Domain (HIGD) proteins to hypoxia plays a crucial role in plant development. However, the research on this gene family in soybean has been lacking. In this study, we aimed to identify and comprehensively analyze soybean *HIGD* genes using the *Glycine max* genome database. As a result, six *GmHIGD* genes were successfully identified, and their phylogeny, gene structures, and putative conserved motifs were analyzed in comparison to *Arabidopsis* and rice. Collinearity analysis indicated that the *HIGD* gene family in soybean has expanded to some extent when compared to *Arabidopsis*. Additionally, the cis-elements in the promoter regions of *GmHIGD* and the transcription factors potentially binding to these regions were identified. All *GmHIGD* genes showed specific responsiveness to submergence and hypoxic stresses. Expression profiling through quantitative real-time PCR revealed that these genes were significantly induced by PEG treatment in root tissue. Co-expressed genes of *GmHIGD* were primarily associated with oxidoreductase and dioxygenase activities, as well as peroxisome function. Notably, one of *GmHIGD* genes, GmHIGD3 was found to be predominantly localized in mitochondria, and its overexpression in *Arabidopsis* led to a significantly reduction in catalase activity compared to wild-type plants. These results bring new insights into the functional role of GmHIGD in terms of subcellular localization and the regulation of oxidoreductase activity.

KEYWORDS

HIGD, *Glycine max*, expression pattern, hypoxia response, mitochondria

1 Introduction

Soybean provides humans with a large amount of protein, essential amino acids, oil, and metabolizable energy (León et al., 2021; Zhou et al., 2021). However, soybean is extremely sensitive to flooding stress during growth and development. Flooding stress can significantly affect growth, grain yield, and seed quality by reducing plant growth, nitrogen fixation, photosynthesis, biomass accumulation, stomatal conductance, and nutrition availability from soil (Githiri et al., 2006; Valliyodan et al., 2017; Ye et al., 2018). Flooding stress can occur in two forms: submergence stress, where the plant organ is completely under water, and waterlogging stress, where the plant's leaves and stems are partially submerged (Nishiuchi et al., 2012). It was estimated that waterlogging can reduce soybean yield by 17–43% during the vegetative growth stage and 50–56% during the reproductive stage (Oosterhuis et al., 1990). Even just two days of waterlogging can reduce soybean production by 27% (Linkemer et al., 1998). Additionally, flooding can increase the risk of plant pathogens and the occurrence of crop disease after floods (Gravot et al., 2016).

The primary and most direct impact of flooding stress is oxygen deprivation. In flooded conditions, the lack of oxygen (O_2) in the water can cause cellular damage and then restrain plant growth. This, in turn, prevents the production of glucose, leading to various metabolic issues. The severity of these negative effects increases with prolonged submergence and high temperatures, which elevate oxygen consumption through plant respiration (Deutsch et al., 2015). Additionally, certain developmental stages of plants, such as seed germination, early growth after germination, and flowering, are particularly sensitive to low O_2 availability (Considine et al., 2017; Le Gac and Laux, 2019). Previous studies have shown that soybean plants are most vulnerable to flooding damage during early growth stages when secondary aerenchyma in roots has not yet formed and could not provide an oxygen pathway under flooded conditions (Shimamura et al., 2003; Valliyodan et al., 2017).

Hypoxia Inducible Gene Domain (HIGD) protein family in mammals consists of five homologs, namely *HIGD–1A*, *HIGD–1B*, *HIGD–1C*, *HIGD–2A* and *HIGD–2B* (Bedo et al., 2004). HIGD1A, initially identified and studied in humans, is a mitochondrial inner membrane protein of approximately 10 kDa that is induced by hypoxia-inducible factor 1 (HIF-1). It interacts with the mitochondrial electron transport chain to reduce oxygen consumption and plays a role in both cell death and survival, depending on the developmental stage and cellular microenvironment (Hayashi et al., 2012; Ameri et al., 2013). It has been reported that HIGD1B acts as an inhibitor to prevent hypoxia–induced mitochondrial fragmentation (Pang et al., 2021). HIGD1C is essential for oxygen sensing in the carotid body and increases the sensitivity of complex IV to hypoxia (Timón-Gómez et al., 2022).

The *HIGD* gene family has been identified in both *Arabidopsis* and rice plants. In rice, there are five *HIGD* genes: *OsHIGD2*, *OsHIGD3* and *OsHIGD5* respond to submergence, hypoxia, and ethylene at different time points, while *OsHIGD1* and *OsHIGD4* exhibit almost undetectable expression under all conditions (Hwang and Choi, 2016). In *Arabidopsis*, *AtHIGD1* expression levels are induced by hypoxia treatment, and overexpression of the *AtHIGD1* gene has been shown to enhance survival rates after hypoxia stress compared to wild-type plants (Hwang et al., 2017). Both OsHIGD2 and AtHIGD1 localize to mitochondria (Hwang and Choi, 2016; Hwang et al., 2017). However, no information is available on the hypoxia response information of the soybean *HIGD* family. During the onset of hypoxia-triggered responses, responsive genes are expected to play key regulatory roles and the detrimental effects of hypoxia can sometimes be counteracted through the induced expression of genes encoding proteins that promote immunity (Hsu et al., 2013). To better understand soybean adaptation to flooding stress and hypoxia environments, we characterized the soybean *HIGD* gene family in terms of sequence, structure, phylogenetic relationships, gene structure, conserved motifs, and chromosomal localization. We also analyzed their expression levels in various tissues and under different stress conditions. Overall, our results provide detailed insights into the soybean *HIGD* family and could facilitate a more comprehensive understanding of the function of *HIGD* genes in soybean.

2 Materials and methods

2.1 Identification of *GmHIGD* genes from the *G.max* genomic sequence and phylogenetic analysis

Soybean predicted proteins were obtained from the Phytozome database (https://phytozome-next.jgi.doe.gov/, v13) (Goodstein et al., 2012). GmHIGD proteins were identified using BLASTp searches against the soybean predicted proteins, with *Oryza sativa* and *Arabidopsis* HIGD proteins as queries. All potential GmHIGD sequences were analyzed to verify the presence of the conserved hypoxia induced protein region (HIG_1_N, PF04588) using Pfam tools (https://pfam.xfam.org/). Sequences lacking the conserved regions were manually removed. A neighbor-joining (NJ) tree was constructed using MEGA 7.0 with 1000 bootstrap replicates, including HIGD proteins from rice and *Arabidopsis*. Gene names from GmHIGD1 to GmHIGD6 were assigned according to their positions in the phylogenetic tree. All protein amino acid sequences that used for phylogenetic analysis were provided in Supplementary File 1.

2.2 Gene structure and conserved motif analysis

Gene structure information was retrieved from the *Glycine max* genome data, and visualized using TBtools software (Chen et al., 2020). The exon-intron structures of *GmHIGD* genes were analyzed by aligning the coding sequences with their corresponding genomic sequences and visualized using the online software GSDS (http://gsds.gao-lab.org/index.php, Hu et al., 2015). The amino acid sequences of GmHIGDs were analyzed using the MEME tool (http://meme-suite.org/index.html) to identify conserved domains

and motifs within each group. The analysis included setting the maximum number of motifs to 5, with a minimum width of 6 and a maximum width of 50 amino acid residues, and an e-value threshold of less than 1x10-^{8}.

2.3 Protein properties, 3-D domain and subcellular localization

The physical and chemical properties of GmHIGD proteins were analyzed using the ProtParam online tool (https://web.expasy.org/protparam/, Gasteiger et al., 2005). Subcellular localization predictions for these GmHIGD proteins were carried out using various tools including the CELLO v.2.5: subCELlular LOcalization predictor (http://cello.life.nctu.edu.tw/) (Yu et al., 2006), WoLF_PSORT tool (http://www.genscript.com/wolfpsort.html), the mGOASVM server (http://bioinfo.eie.polyu.edu.hk/mGoaSvmServer/mGOASVM.html, Plant V2), and the Plant-mPLoc database (http://www.csbio.sjtu.edu.cn/bioinf/plant-multi, Chou and Shen, 2010; Wan et al., 2012). Additionally, the Phyre2 server (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index) was used for homology modelling to predict the three-dimensional (3D) structure of HIGD proteins from soybean, *Arabidopsis*, and rice.

2.4 Chromosomal location analysis and gene duplication

The chromosomal locations of *GmHIGD* genes were determined based on the *Glycine max* genome annotation and visualized using TBtools software (https://github.com/CJ-Chen/TBtools-Manua, Chen et al., 2020). Duplication events in *GmHIGD* genes within the soybean genome were detected using the Multiple Collinearity Scan toolkit (MCScanX) and visualized with the CIRCOS program (Krzywinski et al., 2009; Wang et al., 2012).

2.5 Putative cis-elements in the promoter regions

The 2,000 bp sequences upstream from the translation start codon of all *GmHIGD* genes were obtained from Phytozome v13. Putative cis-acting regulatory elements within these sequences were predicted using the PlantCARE online database (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/, accessed on 22 Feb 2022) (Lescot et al., 2002) and the PLACE website (https://www.dna.affrc.go.jp/PLACE/?action=newplace).

2.6 Co-expression and association genes of GmHIGD analysis

Association genes of *GmHIGDs* were downloaded from the STRING database, which provides functional protein association networks (https://cn.string-db.org, v11.5; Szklarczyk et al., 2021). Co-expression-based gene network analysis was performed on all *GmHIGD* genes using Spearman correlation coefficients to identify relevant genes from RNA-Seq data. Gene selection was based on a co-expression value greater than 0.7. All of these associated genes and co-expressed genes were subjected to Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis.

2.7 Subcellular location of GmHIGD3

Firstly, the GmHIGD3 coding sequence without the stop codon was cloned into vector pHB-35S-mCherry to generate C-terminal mCherry fusions. Then, the obtained pHB-GmHIGD3-mCherry fusion plasmid was transformed into *Escherichia coli* DH5α and verified by sequencing (Sangon, Shanghai, China). Co-expression of the plasmids pK7FWG2-REM1.2-EGFP (membrane protein marker, Huang et al., 2019), pCAMBIA1301-EGFP-AtCAT2 (peroxisome protein marker), pCAMBIA1301-GFP-GLP$_{151}$-P2P3 (plastid protein marker, Li et al., 2023) and pHELLSGATE-GPAT1-EGFP (mitochondrial membrane marker, Jia et al., 2022) was achieved in *N. benthamiana* leaves via *A. tumefaciens* (GV3101 strain)-mediated transformation (Norkunas et al., 2018). After 60 h incubation period, confocal imaging analysis was conducted on Zeiss LSM 880 NLO laser scanning confocal microscope systems.

2.8 Generation and molecular analysis of GmHIGD3-overexpressing plants

GmHIGD3-overexpressing lines were generated in Col-0 background using pHB-GmHIGD3-mCherry vector and the floral dip method (Clough and Bent, 1998). Putative transgenic lines were selected based on hygromycin resistance and PCR analysis. Three T3 transgenic lines were selected for catalase enzyme activity analysis. For the determination of catalase activities, approximately 100 mg fresh leaves of *Arabidopsis* were homogenized in 1 mL of PBS (pH 7.8, 50 mM PBS). The catalase activity assay was performed using the Catalase Assay Kit (BC0200, Solarbio, Beijing, China) in accordance with the manufacturer's instructions. Each experiment was performed with three independent replicates.

2.9 Tissue expression pattern analysis based on RNA sequencing data

The expression levels of *GmHIGD* genes in seven soybean tissues were obtained from Fragments Per Kilobase per Million (FPKM) values at Phytozome v13 (Goodstein et al., 2012). A heatmap of *GmHIGD* genes was constructed using TBtools to visualize expression levels in different tissues based on the FPKM values. Flower tissue was collected from opened flowers grown in the field during the flowing stage. Root, lateral root, root tip, shoot tip, leaf, and stem tissues were collected from 4-week-old plants grown on B&D medium (Libault et al., 2010). The seed stages were determined based on weight range: S1 < 10 mg; S2, 30–50 mg

(storage cells have large central vacuoles); S3, 70–90 mg (storage protein accumulation has begun and vacuole subdivision is occurring); S4, 115–150 mg; S5, 200–250 mg (storage vacuoles are filling); S6, >300 mg (green seeds); S7, >300 mg (yellow seeds); S8, 200–250 mg (fully-mature, yellow and dehydrating seeds); S9 < 150 mg (yellow and fully dehydrated seeds).

2.10 Plant materials, growth conditions and treatments

Soybean Williams 82 seeds were germinated on Petri dishes lined with moist filter paper and then transferred to half-strength MS solution. Seedlings were grown in a growth chamber under a 10-hour photoperiod at temperature of 25°C/22°(day/night) and 50% relative humidity. Plants at the vegetative 1 stage were subsequently transferred to half MS solution containing 15% PEG6000 or 150 mM NaCl for 24 hours and 48 hours, respectively. Roots and the first trifoliolate leaves from 5 plants were collected for *GmHIGD* gene expression analysis. Following collection, samples were immediately frozen in liquid nitrogen and stored at -80°for subsequent total RNA isolation.

For hypoxia treatment, 7-day-old soybean seedling were placed in a sealed container (40 cm x 40 cm x 40 cm) equipped with inlet and outlet valves, and were exposed to hypoxia by flushing N_2 gas (30ml/S) from a nitrogen tank into the container to maintain hypoxia conditions throughout the experiment. Plants were subjected to hypoxia treatment for durations of 2 hours, 4 hours and 8 hours, respectively. Leaves were sampled from 5 different plants for each treatment with three biological replicates utilized to evaluate gene expression patterns.

Nitro blue tetrazolium (NBT, Cat#298-83-9, Solarbio, Beijing, China) staining detected the presence of superoxide. 20-day-old *Arabidopsis* plants was submerged in water for a duration of 12 hours, following which the leaves were soaked in NBT solutions (0.33 mg/ml) for a period of 2 hours. Following this, the leaves underwent decolorization using 95% ethanol in an 80°C-water bath, with the ethanol solution changed every 10 minutes. After complete fading of the green color in the sample, it was examined under a microscope for imaging.

2.11 RNA Isolation, cDNA synthesis, and qRT-PCR

The transcript abundance of all *GmHIGD* genes was investigated using qRT-PCR. Total RNA was extracted using an RNApure Plant Kit (DNase I) (CWBIO, Cat: # CW0559, China) according to the manufacturer's instructions. Approximately 2 μg of total RNA was converted into cDNA using HiScript III RT SuperMix for qPCR (+gDNA wiper) (Vazyme, Cat: # R111-01, China) in a 20 μL reaction volume according to the supplier's instructions. The Bio-Rad CFX ConnectTM Optics Module Real-Time PCR System (Bio-Rad, USA) and ChamQ Universal SYBR qPCR Master Mix (Vazyme, Cat: # Q711, China) were used to performe quantitative RT-PCR. The *Gmactin11* gene served as a reference gene and specific primers for *HIGD* genes were used for qRT-PCR validation. Gene expression data obtained via qRT-PCR were normalized to the expression of *GmActin* gene and the $2^{-\triangle\triangle Ct}$ method was employed to calculate the relative expression of *GmHIGD* genes. Each sample was tested in triplicate and three biological replicates were performed. The primers used for qRT-PCR are given in Supplementary Table S1.

3 Results

3.1 Identification and characterization of *GmHIGD* genes

HIGD amino acid sequences from rice and *Arabidopsis* were used to identify homologs in the soybean genome (*Glycine max* Wm82.a4.v1). Six HIGD homologs, named GmHIGD1 (chromosome 5), GmHIGD2 (chromosome 10), GmHIGD3 (chromosome 20), GmHIGD4 (chromosome 11), GmHIGD5 (chromosome 18) and GmHIGD6 (chromosome 2), were identified in the Wm82 genome assembly. To further address the evolutionary conservation of GmHIGD proteins, a phylogenetic analysis was performed with three *Arabidopsis* (a model dicot plant), and five rice (a model monocot plant) HIGDs using the neighbor-joining method and bootstrap values from 1,000 replicates. The *HIGD* genes were categorized into three groups based on the tree topology and conserved motifs. *GmHIGD4*, *GmHIGD5*, *GmHIGD6*, *OsHIGD3* and *OsHIGD5* were grouped together; *OsHIGD1*, *AtHIGD1* and *GmHIGD1* were in the same group; while the remaining genes (*AtHIGD2*, *AtHIGD3*, *GmHIGD2*, *GmHIGD3* and *OsHIGD2*) formed another group (Figures 1A, B).

The physicochemical properties of the identified GmHIGD protein sequences were evaluated using the ExPASy ProtParam tool (Table 1). The amino acid length ranged from 69 to 100 residues, with molecular weight varying from 7952.24 to 11058.91 kDa. The theoretical pI values ranged from 8.95 to 10.38. GmHIGD1, GmHIGD3, and GmHIGD4 were classified as hydrophilic proteins based on their negative grand average of hydropathy (GRAVY index) values, while the remaining three genes were categorized as hydrophobic proteins due to their positive values. Among the identified GmHIGDs, GmHIGD3 and GmHIGD6 exhibited instability index values below 40, suggesting a more stable nature compared to the others with values above 40.

Analysis of gene structure using the Gene Structure Display Server (GSDS) revealed that, with few exceptions, *GmHIGD* genes share a conserved genomic structure with two or three exons separated by one or two introns (Figure 1C). The structural characteristics of GmHIGD proteins were further plotted based on protein sequence using the MEME motif search tool. The results showed that most of these proteins contained two to three consensus motifs (Figure 1C). Soybean HIGD proteins exhibit high conservation with *Arabidopsis* and rice homologs in motif

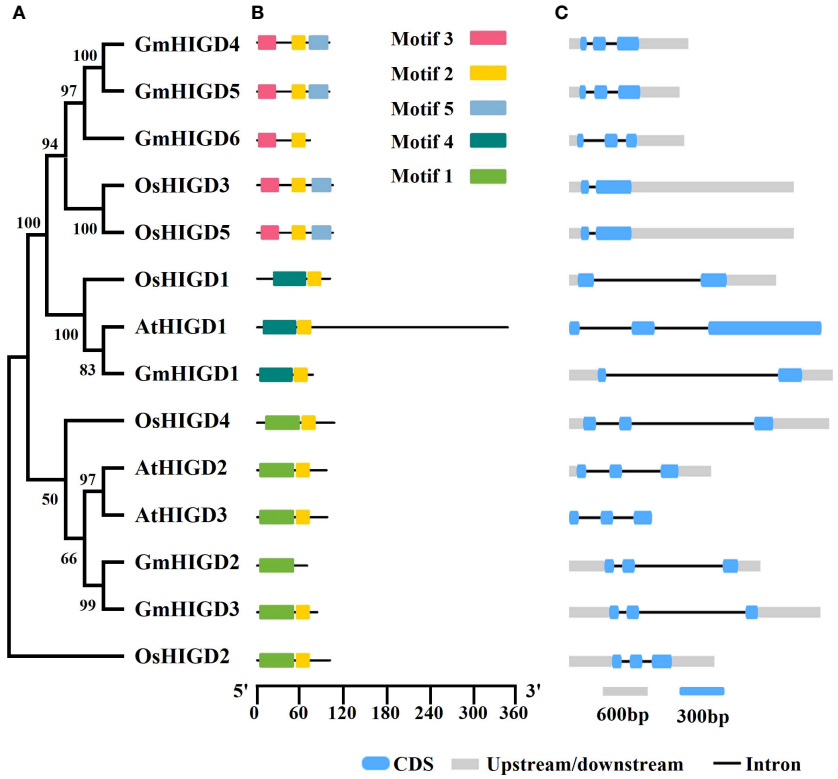


FIGURE 1
Characterization of soybean GmHIGD genes. **(A)** The phylogenetic tree was constructed with MEGA11 using the neighbor-joining (NJ) method with 1000 bootstrap replicates. The percentage bootstrap scores are indicated on the nodes. **(B)** Motif characterization of HIGD proteins. The motifs are displayed in different colored boxes. **(C)** Gene structure of HIGD genes. Gray boxes indicate untranslated 5′-and 3′-regions; blue boxes indicate exons; black lines indicate introns. Scale bar represents gene length.

alignment (Figure 1C). Closely related genes exhibit similar motif compositions, suggesting functional similarities among HIGDs. Motif 2 is present in all HIGDs among *Arabidopsis*, rice, and soybean, except for GmHIGD2. However, GmHIGD2 only has one conserved motif (motif 1), implying potential functional variations within the *GmHIGD* family. The modeling of *GmHIGDs* tertiary structure were predicated through SWISS-MODEL and verified with SAVES v6.0. As shown in Supplementary Figure S1, two α-helix bundles are formed with one β-turn located between the two helixes in all GmHIGDs. The 3D protein structure of HIGDs from *Arabidopsis* and rice were predicated and the results indicated that they all showed the similar structure with that of GmHIGDs (Supplementary Figure S1).

3.2 Chromosomal location and gene collinearity analysis of GmHIGDs

The chromosomal location of *GmHIGDs* was determined through BLASTN searches in the soybean Wm 82 genome database. As shown in Figure 2, all *GmHIGDs* are unevenly distributed on six different chromosomes. Furthermore, gene

TABLE 1 The physicochemical properties of predicted soybean HIGD proteins.

Gene Name	Gene ID	Number of amino acids	Molecular weight	Theoretical pI	Instability index	Grand average of hydropathicity
GmHIGD1	Glyma.05G035600	77	8444.78	9.98	42.93	-0.014
GmHIGD2	Glyma.10G151300	69	8005.4	10.38	40.13	0.038
GmHIGD3	Glyma.20G236900	83	9154.67	10.01	32.86	-0.082
GmHIGD4	Glyma.11G235000	100	11058.91	8.95	51.65	-0.019
GmHIGD5	Glyma.18G022000	100	11044.93	9.23	43.57	0.05
GmHIGD6	Glyma.02G259700	73	7952.24	10.36	27.97	0.086

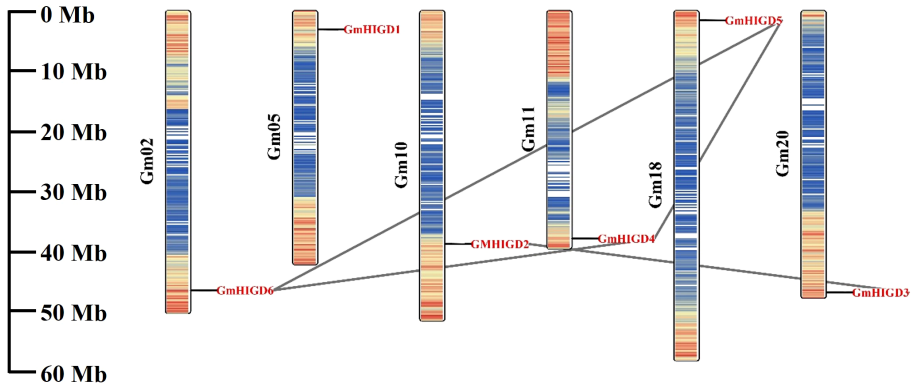


FIGURE 2
Chromosomal location and gene duplication events of GmHIGD1-6 genes. The scale on the left is in megabases (Mb) and the chromosome numbers are indicated at the left of each bar. The segmental duplication genes have been connected by black straight line.

pairs in the *GmHIGD* family were detected. A total of three gene pairs were detected in the *GmHIGD* gene family, with three genes repeatedly participating in gene duplication events (Figure 2).

To explore the evolutionary relationship of *GmHIGD* genes across different species, we constructed nine syntenic soybean maps associated with six dicotyledonous plants *Arabidopsis thaliana*, *Vigna unguiculata*, *Phaseolus vulgaris*, *Lotus japonicus*, *Medicago truncatula*, *Cicer arietinum* (Figure 3) and three monocotyledons *Oryza sativa*, *Zea mays* and *Sorghum bicolor* (Supplementary Figure S2). Notably, the results indicated a higher homology between soybean and other legume species than that between *Oryza sativa*, *Zea mays* and *Sorghum bicolor*. Specifically, two *Arabidopsis* genes (*AtHIGD2* and *AtHIGD3*) were found to be orthologous with *GmHIGD2* and *GmHIGD3*, respectively. *AtHIGD1* was orthologous with soybean gene (Glyma17g091600), although this gene does not encode a HIGD protein. In addition, only one homologous pair was found between soybean and *Sorghum bicolor*. Three *GmHIGDs* (*GmHIGD4*, *GmHIGD5*, *GmHIGD6*) were found to be orthologous with one *ZmHIGD* gene. Four *GmHIGDs* (*GmHIGD2*, *GmHIGD4*, *GmHIGD5*, *GmHIGD6*) were found to be orthologous with three *OsHIGDs*. These results indicate that the *GmHIGD3* gene may be unique to dicotyledons, whereas *GmHIGD1* is conserved specifically in legumes. The remaining *GmHIGDs* show high conservation during evolution between dicotyledonous and monocotyledonous plants.

3.3 Cis-elements analysis of the *GmHIGD* genes promoters and TF identify

To better understand the roles of *GmHIGDs*, we analyzed their promoter regions. The 2 kb promoter regions were extracted from

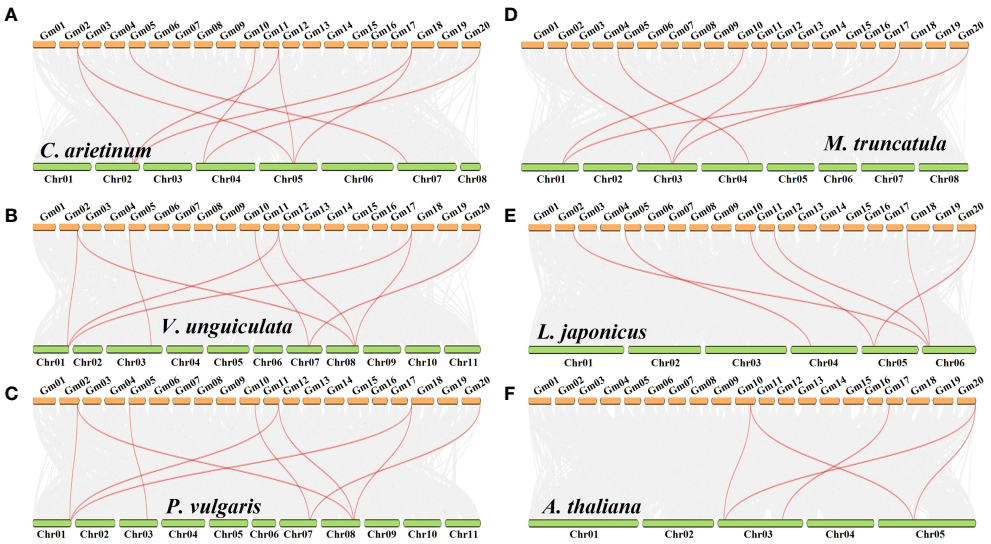


FIGURE 3
Synteny analysis of GmHIGD with six dicotyledonous plants: *Cicer arietinum* **(A)**, *Vigna unguiculata* **(B)**, *Phaseolus vulgaris* **(C)**, *Medicago truncatula* **(D)**, *Lotus japonicus* **(E)** and *Arabidopsis thaliana* **(F)**. Gray lines represent collinear within soybean and other species, while the red line highlights a one-to-one correspondence of homologous regions of *HIGD* gene pairs. The chromosome number is labeled at the top or bottom of each chromosome.

the soybean genome database and submitted to the plantCARE website (Figure 4). As expected, many core promoter elements, such as CAAT-box and TATA-box, were widely distributed in the *GmHIGD* genes. These cis-regulatory element can be categorized into three major groups: abiotic stress response, hormone response and growth and development response. Abiotic stress response elements include anaerobic induction, drought inducibility, defense and stress response elements, like ABRE, MYB, MYC and STRE. Hormone response elements include MeJA, gibberellin, abscisic acid, salicylic acid and auxin response elements. Growth and development response element involve meristem expression, circadian control and cell cycle regulation. These results indicate that the *GmHIGD* gene not only regulates soybean growth and development but also plays an important regulatory role in responding to stress, especially in coping with hypoxia stress.

Furthermore, we predicted transcription factor regulatory networks within the 1 kb promoter sequence using the Plant Transcriptional Regulatory Map. The result revealed that 117 transcription factors (23 families of TFs) might participate in regulating the expression pattern of the *GmHIGDs* family (Supplementary Table S2). For instance, the promoter sequence of *GmHIGD1* contained binding sites for ERF transcription factors, while the promoter sequence of *GmHIGD2* had binding sites for bHLH transcription factors. These results suggest that there is a complex regulation network among the different *GmHIGDs*. The identified transcription factors could potentially be valuable for modifying soybean hypoxia response.

3.4 *GmHIGD* gene expression analysis across different tissues

To characterize the expression patterns of individual *GmHIGDs* at different stages, we used publicly available RNA-seq data for *G. max*. The RPKM values for seed development and vegetative growth tissues were plotted in a hierarchical heatmap (Figure 5). *GmHIGD1*, *GmHIGD2* and *GmHIGD3* were distributed across different tissues, while *GmHIGD2* and *GmHIGD3* had relatively higher levels in the root. On the other hand, *GmHIGD4*, *GmHIGD5* and *GmHIGD6* exhibited greatly expression in root tissues, with *GmHIGD4* and *GmHIGD5* showing elevated expression in the root tip (Figure 5A). Notably, *GmHIGD2* and *GmHIGD6* exhibited root-specific expression patterns, suggesting potential specific roles in root development.

At different seed development stages of soybean, *GmHIGD2*, *GmHIGD3* and *GmHIGD4* showed very high expression levels in

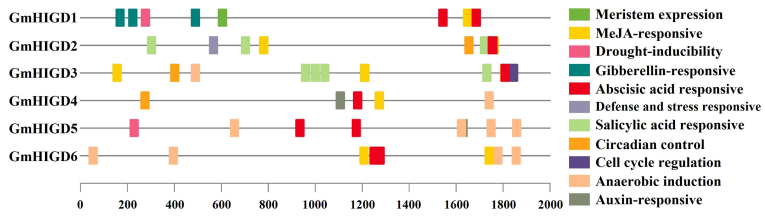


FIGURE 4
Predicted cis-elements in GmHIGD promoter regions. Promoter sequences (–2000 bp) were analyzed by PlantCARE. The upstream length to the translation starts site can be inferred according to the scale. Legend depicting the color of individual cis elements.

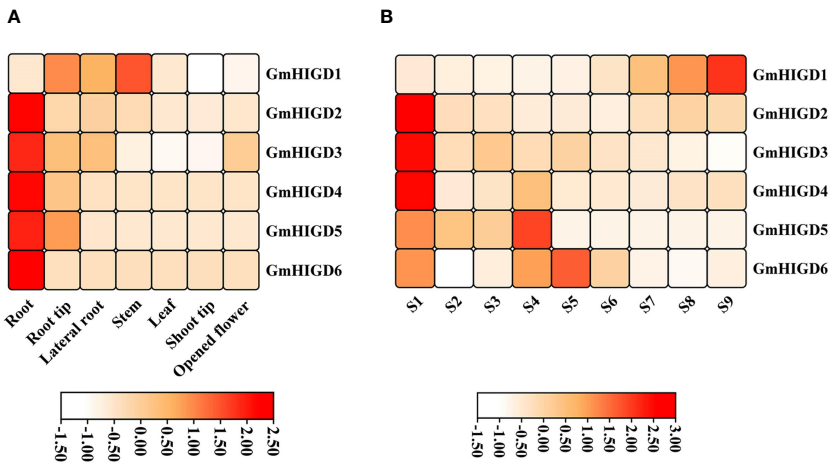


FIGURE 5
Heatmap showing expression pattern of *GmHIGD* genes across different tissues and seed development stages of soybean. **(A)** different tissues **(B)** different seed development stages. The gene expression was used RNA-seq data derived from mean value of three replicates in each tissue and shown on a scale with z-scaling by row.

the early stage of seed development (S1). As for the S4-S5 seed development stage, the expression levels of *GmHIGD5* and *GmHIGD6* were higher than those in other stages; *GmHIGD1* displayed significant expression only in S7-S9 during the late stage of seed development (Figure 5B). These results indicate that different *GmHIGD* genes may play distinct functions at different stages of seed development.

3.5 Expression patterns of *GmHIGD* genes under flooding and hypoxia stresses

To evaluate the expression profiling of *GmHIGD* genes under flooding and hypoxia conditions, we used RNA-seq data from three studies. According to the results reported by Tamang et al. (2021), no differentially expression pattern was found in leaf tissues after 1-3 days submergence or 1 day of recovery following 3 days of submergence. However, all *GmHIGD* genes showed high expression levels in root tissues under treatment. In particular, *GmHIGD2* and *GmHIGD3* were highly induced after 2- or 3-days submergence treatment (Figure 6A). Lin et al. (2019) reported all *GmHIGD* genes showed relatively high expression levels under submergence treatment in root tissue of the Qihuang34 variety. Especially, *GmHIGD4* and *GmHIGD6* were highly induced after 3 h or 6 h compared to non-treatment (Figure 6B). In addition, we further investigated the expression of *GmHIGD* genes in root tissue of flood-tolerant Embrapa 45 and flood-sensitive BR 4 soybean cultivars under hypoxic stress (Nakayama et al., 2017). As shown in Figure 6C, the expression of five *GmHIGD* genes was enhanced in both cultivars under hypoxia induction at different time points. These results showed that all *GmHIGD* genes are specifically responsive to submergence in root tissues, indicating a positive role of the *GmHIGD* family in the soybean's response to submergence processes.

We further analyzed the expression of six *GmHIGD* genes in Wm82 under hypoxia conditions with N_2 treatment using the qPCR method (Figures 6D–H). Due to high homology in gene sequences between *GmHIGD4* and *GmHIGD5*, the designed primers amplified both genes simultaneously in this study. The results showed that the transcripts of *GmHIGD4/5* were largely accumulated (30 to 1100-fold) after 2-8 hours of hypoxia treatment. *GmHIGD6* was induced over 20-fold after 8 hours of hypoxia treatment in soybean seedlings (Figures 6D–H). These results suggest that the expression of the *GmHIGD* gene family is regulated by flooding and hypoxic stress.

3.6 Expression patterns of *GmHIGD* genes under drought and salt condition

To understand the effects of drought and salt stress on *GmHIGD* gene expression, we detected the transcript abundance of *GmHIGDs* in soybean roots and leaves by qPCR after 24 hours and 48 hours of PEG or NaCl treatment, respectively. As shown in Figure 7, *GmHIGD1*, *GmHIGD3* and *GmHIGD4/5* genes showed significantly upregulaton in both leaves and roots after PEG treatment (Figures 7A, C, D). *GmHIGD2* transcript levels were up-regulated in root tissues under PEG or NaCl treatment (Figure 7B). The expression levels of *GmHIGD4/5* and *GmHIGD6* was significantly and highly increased in root tissues under PEG or NaCl treatments, especially under PEG treatment (Figures 7D, E). In contrast, the expression level of *GmHIGD6* in leaves was downregulated, with no significant changes found in other *GmHIGD* genes after NaCl treatment in leaves. These results suggest that *GmHIGDs* may participate in the response and resistance of soybean to drought and NaCl stress.

3.7 Function prediction of *HIGD* genes

To further explore the function of *GmHIGD*, six *GmHIGDs* genes were selected as 'guide genes' to identify co-expressed genes using expression data from the Phytozome database with a Pearson's correlation coefficient (PCC) threshold of 0.7 (Aoki et al., 2007). Meanwhile, the STRING database was used to build an interaction network between GmHIGDs proteins and other soybean proteins, focusing on genes with a trusted value (medium confidence, 0.4). Finally, 319 genes exhibiting closely correlated expression patterns to or interaction with GmHIGDs were identified. GO annotation analysis of these 319 gene revealed that their involvement in a range of molecular functions, with enriched annotations predominantly in oxidoreductase activity, dioxygenase activity, cation transmembrane transporter activity, transcription factor or translation factor activity (Supplementary Table S3). In addition, the KEGG pathway analysis indicated enrichment in amino acid metabolism, carbohydrate metabolism, energy metabolism and peroxisome pathways (Supplementary Figure S3).

3.8 Characterization of the subcellular location and function of GmHIGD3

The subcellular localization of GmHIGDs predicated by different online tools may be located in diverse organelles, such as mitochondria, chloroplast, peroxisome, nucleus or vacuole (Supplementary Table S4). We further validate the subcellular localization of GmHIGD3 *in vivo* by using overexpressed transgenic *Arabidopsis* seedlings. Fluorescence analysis of GmHIGD3-mCherry expression revealed small punctate structures in root cells, which appeared to be localized to mitochondria (Figures 8A–C). The localization of GmHIGD3-mCherry was further confirmed through co-localization experiments with known organelle markers in tobacco leaf infiltration studies. As shown in Figure 8D–O, there was a partial overlap in the colocalization of GmHIGD3-mCherry fusion protein and REM1.2 (membrane localized protein), or AtCAT2 (peroxisomal protein), or GLP_{151}-P2P3 (plastid localized protein). By contrast, significant co-localization was observed with the mitochondrial marker GPAT1-EGFP when co-expressed (Figures 8P–S). These results strongly suggest that GmHIGD3 is predominantly localized in mitochondria.

10.3389/fpls.2024.1403841

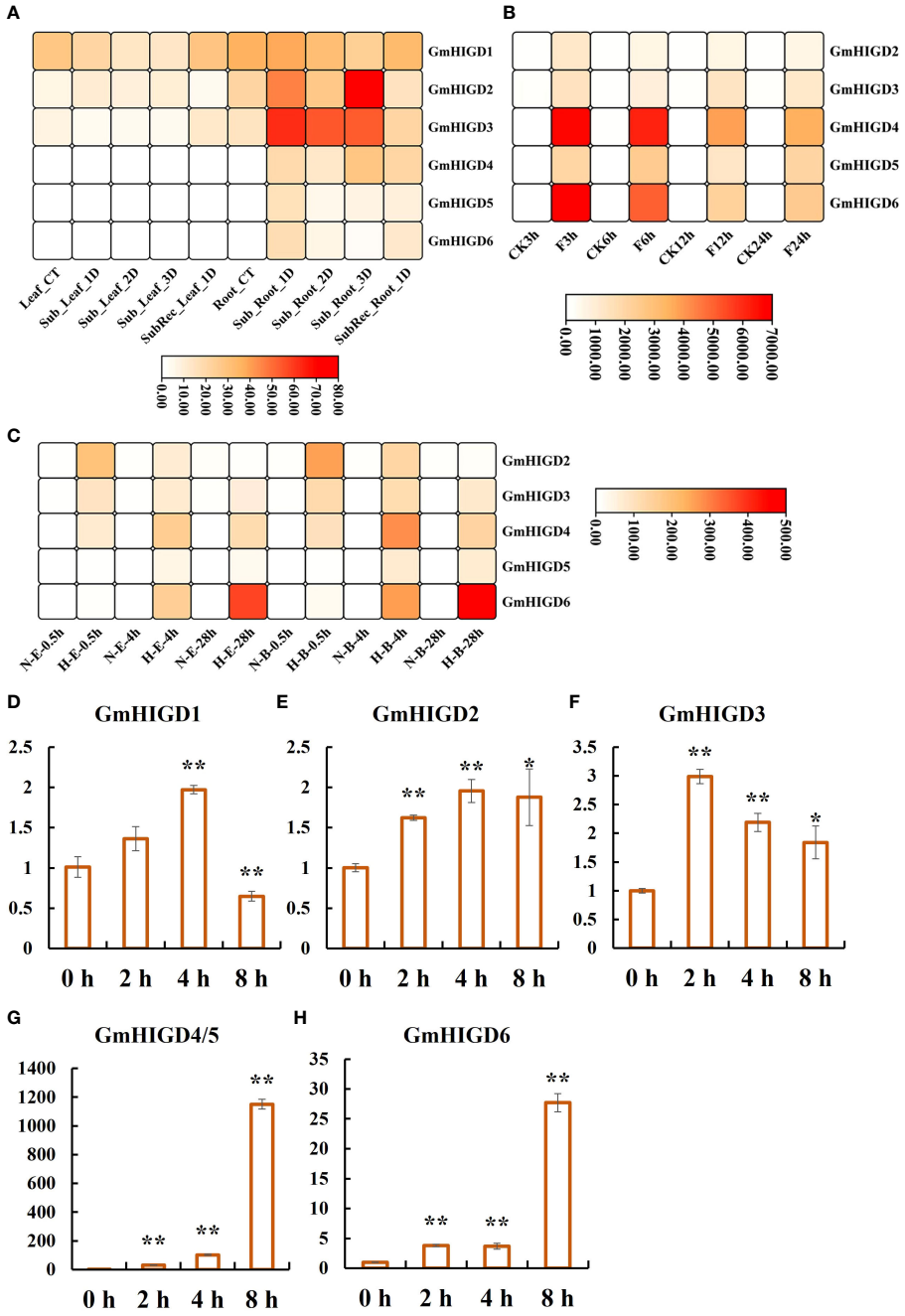


FIGURE 6
Expression patterns of *GmHIGDs* genes under waterlogging and hypoxic stress. **(A)**. Expression patterns of *GmHIGDs* genes in Williams 82 after 1, 2, and 3 days of flooding and 1 day of recovery after 3 days of flooding, Leaf: leaf tissue, Root: root, CT: control, Sub: flooding, Sub Rec: recovery after 3 days of flooding; **(B)**. Expression patterns of *GmHIGDs* gene in roots of Qihuang 34 under water flooding for 3, 6, 12, and 24 hours, CK: control, F: water flooding; **(C)**. Expression pattern of *GmHIGDs* gene in roots of Embrapa 45 and BR 4 under hypoxia for 0.5h, 4h, and 28h, N: non hypoxia, H: hypoxia; E: Embrapa 45, B: BR 4. **(D–H)** Expression analysis of *GmHIGD* genes by quantitative real-time PCR in soybean seedlings under hypoxia treatment for 0h, 2h, 4h and 8h, respectively. *, $p < 0.05$; **, $p < 0.01$.

The transgenic *Arabidopsis* plants overexpressing *GmHIGD3* were generated to investigate its function further. Since GmHIGD3 is primarily located in mitochondria and its co-expressed genes are mainly involved in oxidoreductase activity, the catalase enzyme activity was compared between the transgenic lines and wide type. As shown in Figure 9A, the catalase activity in the overexpressed *GmHIGD3* transgenic lines is significantly lower than that in the wild type, suggesting that GmHIGD3 plays an important role in regulating oxidoreductase activity. Furthermore, the accumulation of superoxide was detected using NBT staining. Increased staining

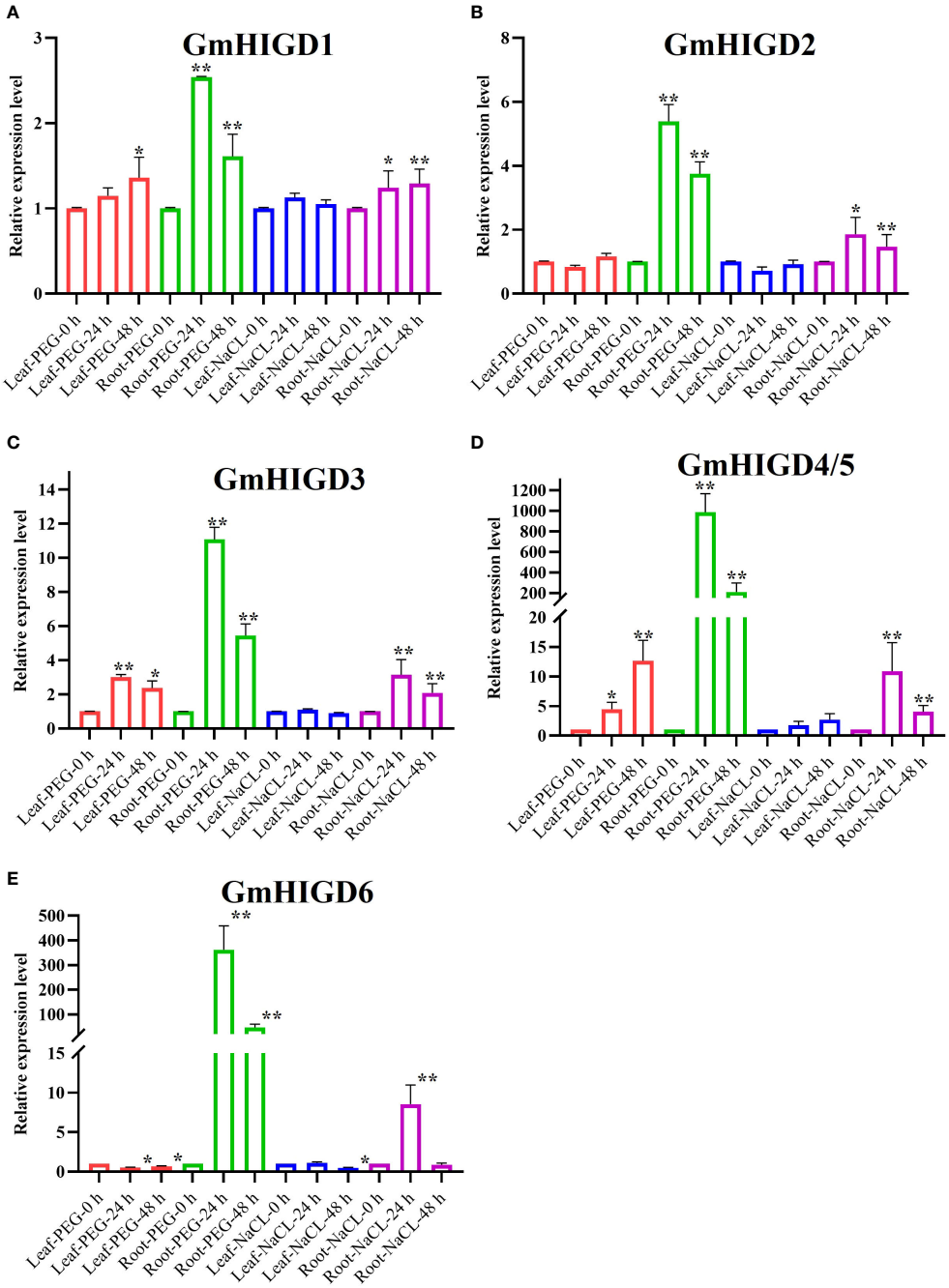


FIGURE 7
Expression profiles of *HIGD* genes, measured by real-time qRT-PCR in Williams 82 ecotype as compared with PEG and NaCl treatments. **(A)** The relative expression levels of *GmHIGD1* gene. **(B)** The relative expression levels of *GmHIGD2* gene. **(C)** The relative expression levels of *GmHIGD3* gene. **(D)** The relative expression levels of *GmHIGD4* gene. **(E)** The relative expression levels of *GmHIGD5* gene. The relative gene expression levels were calculated relative to 0 h and using $2^{-\triangle\triangle CT}$ method. The data shown are the mean values SE of three replicates. * and ** indicates that there are significant differences at 5%, 1% level respectively relative to controls.

was observed in the GmHIGD3-overexpressing transgenic lines compared with wild type under normal and submerged conditions, whereas no significant difference was observed between normal and submerged conditions in GmHIGD3-overexpressing transgenic lines (Figures 9B–I). This result indicates that the downregulated catalase enzyme activity in the GmHIGD3-overexpressing transgenic lines leads to the accumulation of reactive oxygen species (ROS).

4 Discussion

A comparison of all known mammalian and plant *HIGD* genes reveals that the number of HIGD protein is relatively conserved, typically ranging from 3 to 5. As such, the *HIGD* family is not a large or plant-specific group. Members of the *HIGD* family has not been extensively studied at a genome-wide level in plants. In this

10.3389/fpls.2024.1403841

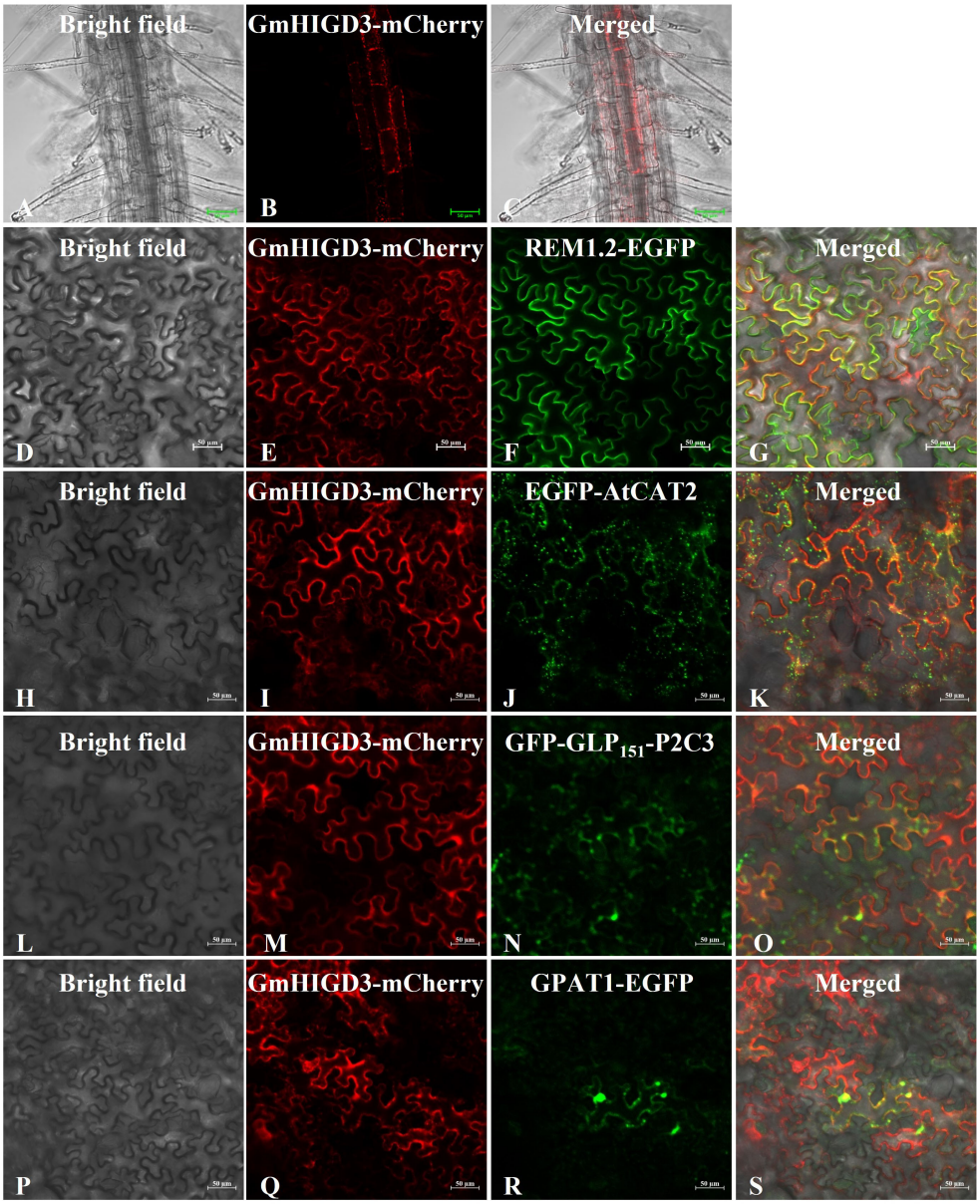


FIGURE 8
Subcellular localization of GmHIGD3 by confocal laser scanning microscopy. **(A–C)** Subcellular localization of GmHIGD3-mCherry in *Arabidopsis* root. Epifluorescence images of tobacco leaf cells infiltrated with Agrobacterium harboring the different fusion constructs; **(D–G)** GmHIGD3-mCherry and REM1.2-EGFP; **(H–K)** GmHIGD3-mCherry and AtCAT2-EGFP; **(L–O)** GmHIGD3-mCherry and GLP_{151}-P2P3-GFP; **(P–S)** GmHIGD3-mCherry and GPAT1-EGFP.

study, we focuses on the identification and characterization of six homologous soybean *HIGD* genes. Furthermore, more predicted orthologous relationships were found in dicots compared to monocots, indicating that their functions may have diverged throughout evolution. Conserved domain and motifs analysis indicated that GmHIGD1 shares conserved domains with AtHIGD1. Proteins within the same subgroup exhibit similar motifs, which might be related to their specific functions.

The expression of *HIGD2A* is dependent on oxygen levels, glucose concentration, and cell cycle progression in human and animals. While the potential roles of the *HIGD* gene family in response to biotic and abiotic stresses are recognized, there is limited research on these genes in plant species. In rice and *Arabidopsis*, the expression patterns of *HIGD* genes change under stress conditions. Hwang and Choi (2016) studied the expression patterns of *OsHIGDs* in rice and found that *OsHIGD2* exhibited a significant response to submergence and hypoxia, with a slight response to ethylene. *OsHIGD3* and *OsHIGD5* were slightly induced by submergence and hypoxia. Except for *OsHIGD2*, no other *OsHIGDs* displayed differential expression in response to ethylene, suggesting that *OsHIGD2* might be the most active member of the *OsHIGD* gene family. Similarly, *AtHIGD1* was upregulated at 8, 16 and 24 hours after hypoxia treatment (Hwang et al., 2017). In our current study, we found that all

10.3389/fpls.2024.1403841

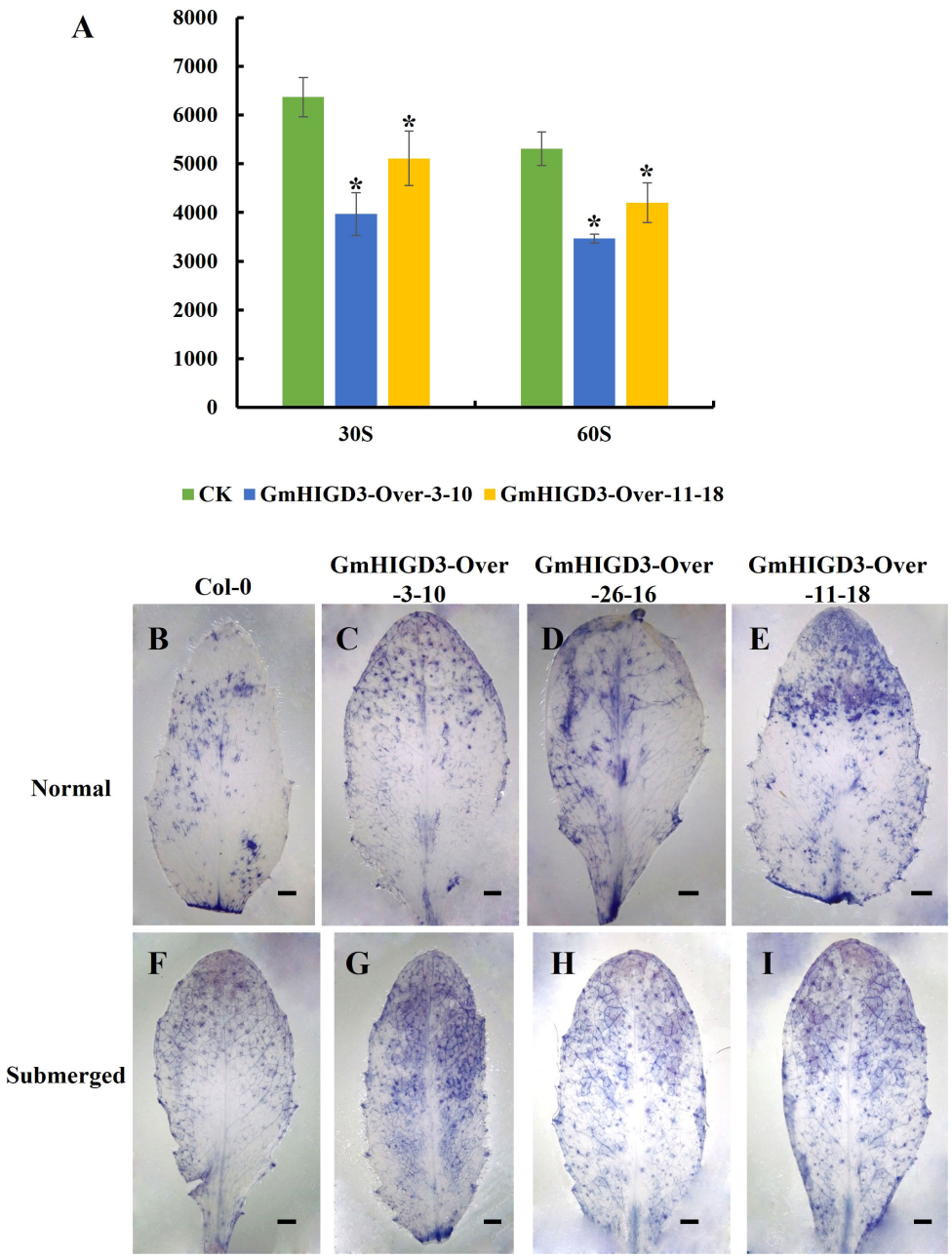


FIGURE 9
The evaluation of catalase activity and ROS content in GmHIGD3 overexpressing transgenic lines and wild type. **(A)** The catalase activity was inhibited in GmHIGD3 overexpressing compared with wild type. * Indicates a significant difference at the 1% level compared to the non-transgenic control. **(B–I)** Superoxide contents were detected by 0.33 mg/mL of NBT staining. Experiments were repeated three times with similar results. Bars represent 1 mm.

GmHIGDs respond to abiotic stress, especially to hypoxia. Among them, the expression level of *GmHIGD4/5* sharply increased after 8 hours of hypoxia induction, followed by *GmHIGD6*. Furthermore, we examined the expression of *GmHIGD* gene family in root and leaf tissues at different time points after PEG treatment. The expression levels of *GmHIGD4/5* and *GmHIGD6* in roots were significantly higher than in the control under PEG stress conditions, suggesting that *GmHIGD4/5* and *GmHIGD6* might be the most active members of the *GmHIGD* gene family in response to abiotic stress. At the same time, *GmHIGDs* were highly expressed in soybean roots. Given that roots play an important role in water and nutrients absorption, the *GmHIGD* gene family may play an vital role in the development of soybean root systems under stress conditions.

To clarify the functional positions of GmHIGDs in plant cells, an online tool was initially used to predict that GmHIGD2 and GmHIGD3 are located in mitochondria. Furthermore, our finding indicates that GmHIGD2 and GmHIGD3 are not only located in

mitochondria, but also in the cell membrane by tobacco epidermal cell experiment. It has been reported that HIGD1A and HIGD2A were embedded in the mitochondrial inner membrane in mammals (Strogolova et al., 2012). In addition to their mitochondrial location, studies have shown that the localization of HIGD proteins may change in response to certain stressors. For example, HIGD1A has been found to relocate to the nucleus during apoptosis triggered by severe metabolic or DNA damage stressors (Ameri et al., 2013 & Ameri et al., 2015), and HIGD2A has been observed in the nucleus even under physiological conditions (Salazar et al., 2019). López et al. (2018) also reported that HIGD1A migrates from the cytoplasmic pool to the nucleus in conditions like ischemic heart disease, cancer, and ischemic encephalopathy. Further experiments are needed to investigate whether the localization of GmHIGD proteins changes in response to environmental stimuli such as hypoxia or drought.

Recent studies in mammals have provided convincing evidence of a strong correlation between *HIGD* and oxygen consumption, ROS production, and tumor growth (Shahrzad et al., 2007; Ameri et al., 2015; Timón-Gómez et al., 2020). While there have been numerous studies on the function of the *HIIGD* gene family in mammals, research on its role in plants is limited. Currently, only the function of AtHIGD1 in *Arabidopsis thaliana* has been preliminarily studied, and it participates in the plant's hypoxia response. With global climate change leading to more frequent flooding and subsequent hypoxia in plant, decreased oxygen levels can severely affect mitochondrial energy generation, causing intense damage (Nakamura and Noguchi, 2020; León et al., 2021; Barreto et al., 2022; Jethva et al., 2022). In hypoxic conditions, cell acidification and ROS accumulation negatively affect overall plant growth. The transition of hypoxia–reoxygenation responses can cause excess generation of ROS, potentially leading to cellular damage (Hebelstrup and Møller, 2015; Turkan, 2018). At the same time, the degradation of ROS via activation of antioxidant mechanism is equally important to prevent cell damage and maintain cellular homeostasis (Lee et al., 2020; Pucciariello and Perata, 2021). Moreover, due to enhanced denitrification under anaerobic conditions, flooding reduces the availability of soil nitrogen, leading to crop loss (Sjøgaard et al., 2018; Yu et al., 2020).

Co-expression analysis is usually used to identify functional factors participating in specific biological processes (Fu and Xue, 2010). In this study, we found that *GmHIGD* genes may be associated with oxidoreductase activity, dioxygenase activity, and cation transmembrane transporter activity based on co-expression analysis. This suggests taht *GmHIGD* genes could contribute to oxidation-reduction processes, cellular nitrogen compound biosynthetic processes, organic cyclic compound biosynthetic processes and heterocycle biosynthetic processes. Additional functional characterization of GmHIGDs and exploration of the transcriptional network related to hypoxia response need to be carried out.

5 Conclusions

In soybean, six *HIGD* genes were identified and their protein physicochemical properties were investigated. Their chromosome location, gene structure, promoter cis-elements, conserved motif, evolutionary relationships and subcellular location were analyzed. The results indicated that *GmHIGD* genes are highly conserved, with five of them exhibiting higher expression levels in root tissue compared to other tissues. Additionally, all *GmHIGDs* responded to hypoxia, flooding, drought and salt stresses at different time points. Functional analysis of *GmHIGD3* illustrated that it is localized to mitochondria and decreased the catalase activity when over-expressed. Co-expression gene enrichment analysis illustrated the biological processes involved in somatic embryogenesis. These compressive analysis of the soybean *HIGD* gene family provides useful insights for further study of the function of *GmHIGD* genes in hypoxia tolerance.

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.

Ethics statement

The plant material used in this study was Williams 82, which was planted in the growth chamber of Yangzhou University, Yangzhou, China, and no permits are required for the collection of plant samples. This study did not require ethical approval or consent as did not involve any endangered or protected species.

Author contributions

XG: Data curation, Methodology, Writing – original draft. LD: Formal analysis, Investigation, Writing – original draft. TZ: Investigation, Writing – original draft. CY: Investigation, Validation, Writing – original draft. JZ: Investigation, Methodology, Writing – original draft. BG: Software, Supervision, Writing – review & editing. HC: Supervision, Writing – original draft. QZ: Supervision, Writing – review & editing. LS: Funding acquisition, Resources, Visualization, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was supported by the Natural Science Foundation of Jiangsu Higher Education Institutions of China (23KJA210003), the open Project Program of Joint International Research Laboratory of Agriculture and Agri-Product Safety, the Ministry of Education of China, Yangzhou University (JILAR-KF202202), and the Zhongshan Biological Breeding Laboratory (ZSBBL-KY2023-03).

Acknowledgments

We gratefully acknowledge professor Xu Chen (Haixia Institute of Science and Technology, Fujian Agriculture and Forestry University) for providing plasmid pK7FWG2-REM1.2-EGFP.

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.2024.1403841/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Predicted three-dimensional structures of the GmHIGD, AtHIGD and OsHIGD protein sequences. Models were constructed using SWISS-MODEL.

SUPPLEMENTARY FIGURE 2
Synteny analysis of GmHIGD with three monocotyledonous species (*O. sativa*, *Z. mays* and *S. bicolor*) by MCscan. Gray lines represent collinear within soybean and other species, while the red line highlights a one-to-one correspondence of homologous regions of HIGD gene pairs. The chromosome number is labeled at the top or bottom of each chromosome.

SUPPLEMENTARY FIGURE 3
Enrichment analysis of co-expressed genes of GmHIGDs.

References

Ameri, K., Jahangiri, A., Rajah, A. M., Tormos, K. V., Nagarajan, R., Pekmezci, M., et al. (2015). HIGD1A regulates oxygen consumption, ROS production, and AMPK activity during glucose deprivation to modulate cell survival and tumor growth. *Cell Rep.* 10, 891–899. doi: 10.1016/j.celrep.2015.01.020

Ameri, K., Rajah, A. M., Nguyen, V., Sanders, T. A., Jahangiri, A., DeLay, M., et al. (2013). Nuclear localization of the mitochondrial factor HIGD1A during metabolic stress. *PloS One* 8, e62758. doi: 10.1371/journal.pone.0062758

Aoki, K., Ogata, Y., and Shibata, D. (2007). Approaches for extracting practical information from gene co-expression networks in plant biology. *Plant Cell Physiol.* 48, 381–390. doi: 10.1093/pcp/pcm013

Barreto, P., Koltun, A., Nonato, J., Yassitepe, J., Maia, I., and Arruda, P. (2022). Metabolism and signaling of plant mitochondria in adaptation to environmental stresses. *Int. J. Mol. Sci.* 23, 11176. doi: 10.3390/ijms231911176

Bedo, G., Vargas, M., Ferreiro, M.-J., Chalar, C., and Agrati, D. (2004). Characterization of hypoxia induced gene 1: expression during rat central nervous system maturation and evidence of antisense RNA expression. *Int. J. Dev. Biol.* 49, 431–436. doi: 10.1387/ijdb.041901gb

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

Chou, K.-C., and Shen, H.-B. (2010). Plant-mPLoc: a top-down strategy to augment the power for predicting plant protein subcellular localization. *PloS One* 5, e11335. doi: 10.1371/journal.pone.0011335

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

Considine, M. J., Diaz-Vivancos, P., Kerchev, P., Signorelli, S., Agudelo-Romero, P., Gibbs, D. J., et al. (2017). Learning to breathe: developmental phase transitions in oxygen status. *Trends Plant Sci.* 22, 140–153. doi: 10.1016/j.tplants.2016.11.013

Deutsch, C., Ferrel, A., Seibel, B., Pörtner, H.-O., and Huey, R. B. (2015). Climate change tightens a metabolic constraint on marine habitats. *Science* 348, 1132–1135. doi: 10.1126/science.aaa1605

Fu, F.-F., and Xue, H.-W. (2010). Coexpression analysis identifies Rice Starch Regulator1, a rice AP2/EREBP family transcription factor, as a novel rice starch biosynthesis regulator. *Plant Physiol.* 154, 927–938. doi: 10.1104/pp.110.159517

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., and Appel, R. D. (2005). *Protein Identification and Analysis Tools on the ExPASy Server*. In: Walker, J. M. (eds) The Proteomics Protocols Handbook. Springer Protocols Handbooks. (Humana Press). doi: 10.1385/1-59259-890-0:571

Githiri, S. M., Watanabe, S., Harada, K., and Takahashi, R. (2006). QTL analysis of flooding tolerance in soybean at an early vegetative growth stage. *Plant Breed.* 125, 613–618. doi: 10.1111/j.1439-0523.2006.01291.x

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

Gravot, A., Richard, G., Lime, T., Lemarié, S., Jubault, M., Lariagon, C., et al. (2016). Hypoxia response in Arabidopsis roots infected by Plasmodiophora brassicae supports the development of clubroot. *BMC Plant Biol.* 16, 1–10. doi: 10.1186/s12870-016-0941-y

Hayashi, H., Nakagami, H., Takeichi, M., Shimamura, M., Koibuchi, N., Oiki, E., et al. (2012). HIG1, a novel regulator of mitochondrial γ-secretase, maintains normal mitochondrial function. *FASEB J.* 26, 2306–2317. doi: 10.1096/fj.11-196063

Hebelstrup, K. H., and Møller, I. M. (2015). Mitochondrial signaling in plants under hypoxia: use of reactive oxygen species (ROS) and reactive nitrogen species (RNS). In: Gupta, K., and Igamberdiev, A. (eds) *Reactive Oxygen and Nitrogen Species Signaling and Communication in Plants. Signaling and Communication in Plants*, vol 23, Springer, Cham. p. 63–77. doi: 10.1007/978-3-319-10079-1_4

Hsu, F.-C., Chou, M.-Y., Chou, S.-J., Li, Y.-R., Peng, H.-P., and Shih, M.-C. (2013). Submergence confers immunity mediated by the WRKY22 transcription factor in Arabidopsis. *Plant Cell* 25, 2699–2713. doi: 10.1105/tpc.113.114447

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., 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.* 116, 21274–21284. doi: 10.1073/pnas.1911892116

Hwang, S.-T., and Choi, D. (2016). A novel rice protein family of OsHIGDs may be involved in early signalling of hypoxia-promoted stem growth in deepwater rice. *Plant Cell Rep.* 35, 2021–2031. doi: 10.1007/s00299-016-2013-z

Hwang, S.-T., Li, H., Alavilli, H., and Lee, B.-h. (2017). Choi D: Molecular and physiological characterization of AtHIGD1 in Arabidopsis. *Biochem. Biophys. Res. Commun.* 487, 881–886. doi: 10.1016/j.bbrc.2017.04.146

Jethva, J., Schmidt, R. R., Sauter, M., and Selinski, J. (2022). Try or die: dynamics of plant respiration and how to survive low oxygen conditions. *Plants* 11, 205. doi: 10.3390/plants11020205

Jia, Q., Bai, Y., Xu, H., Liu, Q., Li, W., Li, T., et al. (2022). Mitochondrial GPAT-derived LPA controls auxin-dependent embryonic and postembryonic development. *Proc. Natl. Acad. Sci.* 119, e2212881119. doi: 10.1073/pnas.2212881119

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., et al. (2009). Circos: an information aesthetic for comparative genomics. *Genome Res.* 19, 1639–1645. doi: 10.1101/gr.092759.109

Le Gac, A.-L., and Laux, T. (2019). Hypoxia is a developmental regulator in plant meristems. *Mol. Plant* 12, 1422–1424. doi: 10.1016/j.molp.2019.10.004

Lee, P., Chandel, N. S., and Simon, M. C. (2020). Cellular adaptation to hypoxia through hypoxia inducible factors and beyond. *Nat. Rev. Mol. Cell Biol.* 21, 268–283. doi: 10.1038/s41580-020-0227-y

León, J., Castillo, M. C., and Gayubas, B. (2021). The hypoxia–reoxygenation stress in plants. *J. Exp. Bot.* 72, 5841–5856. doi: 10.1093/jxb/eraa591

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

Li, T., Xiao, X., Liu, Q., Li, W., Li, L., Zhang, W., et al. (2023). Dynamic responses of PA to environmental stimuli imaged by a genetically encoded mobilizable fluorescent sensor. *Plant Commun.* 4, 100500. doi: 10.1016/j.xplc.2022.100500

Libault, M., Farmer, A., Brechenmacher, L., Drnevich, J., Langley, R. J., Bilgin, D. D., et al. (2010). Complete transcriptome of the soybean root hair cell, a single-cell model, and its alteration in response to Bradyrhizobium japonicum infection. *Plant Physiol.* 152, 541–552. doi: 10.1104/pp.109.148379

Lin, Y., Li, W., Zhang, Y., Xia, C., Liu, Y., Wang, C., et al. (2019). Identification of genes/proteins related to submergence tolerance by transcriptome and proteome analyses in soybean. *Sci. Rep.* 9, 14688. doi: 10.1038/s41598-019-50757-1

Linkemer, G., Board, J. E., and Musgrave, M. E. (1998). Waterlogging effects on growth and yield components in late-planted soybean. *Crop Sci.* 38, 1576–1584. doi: 10.2135/cropsci1998.0011183X003800060028x

López, L., Zuluaga, M. J., Lagos, P., Agrati, D., and Bedó, G. (2018). The expression of Hypoxia-Induced Gene 1 (Higd1a) in the central nervous system of male and female rats differs according to age. *J. Mol. Neurosci.* 66, 462–473. doi: 10.1007/s12031-018-1195-y

Nakamura, M., and Noguchi, K. (2020). Tolerant mechanisms to O 2 deficiency under submergence conditions in plants. *J. Plant Res.* 133, 343–371. doi: 10.1007/s10265-020-01176-1

Nakayama, T. J., Rodrigues, F. A., Neumaier, N., Marcolino-Gomes, J., Molinari, H. B., Santiago, T. R., et al. (2017). Insights into soybean transcriptome reconfiguration under hypoxic stress: Functional, regulatory, structural, and compositional characterization. *PloS One* 12, e0187920. doi: 10.1371/journal.pone.0187920

Nishiuchi, S., Yamauchi, T., Takahashi, H., Kotula, L., and Nakazono, M. (2012). Mechanisms for coping with submergence and waterlogging in rice. *Rice* 5, 1–14. doi: 10.1186/1939-8433-5-2

Norkunas, K., Harding, R., Dale, J., and Dugdale, B. (2018). Improving agroinfiltration-based transient gene expression in Nicotiana benthamiana. *Plant Methods* 14, 1–14. doi: 10.1186/s13007-018-0343-2

Oosterhuis, D., Scott, H., Hampton, R., and Wullschleger, S. (1990). Physiological responses of two soybean [Glycine max (L.) Merr] cultivars to short-term flooding. *Environ. Exp. Bot.* 30, 85–92. doi: 10.1016/0098-8472(90)90012-S

Pang, Y., Zhu, Z., Wen, Z., Lu, J., Lin, H., Tang, M., et al. (2021). HIGD–1B inhibits hypoxia–induced mitochondrial fragmentation by regulating OPA1 cleavage in cardiomyocytes. *Mol. Med. Rep.* 24, 1–11. doi: 10.3892/mmr

Pucciariello, C., and Perata, P. (2021). The oxidative paradox in low oxygen stress in plants. *Antioxidants* 10, 332. doi: 10.3390/antiox10020332

Salazar, C., Elorza, A. A., Cofre, G., Ruiz-Hincapie, P., Shirihai, O., and Ruiz, L. M. (2019). The OXPHOS supercomplex assembly factor HIG2A responds to changes in energetic metabolism and cell cycle. *J. Cell. Physiol.* 234, 17405–17419. doi: 10.1002/jcp.28362

Shahrzad, S., Bertrand, K., Minhas, K., and Coomber, B. (2007). Induction of DNA hypomethylation by tumor hypoxia. *Epigenetics* 2, 119–125. doi: 10.4161/epi.2.2.4613

Shimamura, S., Mochizuki, T., Nada, Y., and Fukuyama, M. (2003). Formation and function of secondary aerenchyma in hypocotyl, roots and nodules of soybean (Glycine max) under flooded conditions. *Plant Soil* 251, 351–359. doi: 10.1023/A:1023036720537

Sjøgaard, K. S., Valdemarsen, T. B., and Treusch, A. H. (2018). Responses of an agricultural soil microbiome to flooding with seawater after managed coastal realignment. *Microorganisms* 6, 12. doi: 10.3390/microorganisms6010012

Strogolova, V., Furness, A., Robb-McGrath, M., Garlich, J., and Stuart, R. A. (2012). Rcf1 and Rcf2, members of the hypoxia-induced gene 1 protein family, are critical components of the mitochondrial cytochrome bc 1-cytochrome c oxidase supercomplex. *Mol. Cell. Biol.* 32, 1363–1373. doi: 10.1128/MCB.06369-11

Szklarczyk, D., Gable, A. L., Nastou, K. C., Lyon, D., Kirsch, R., Pyysalo, S., et al. (2021). The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* 49, D605–D612. doi: 10.1093/nar/gkab835

Tamang, B. G., Li, S., Rajasundaram, D., Lamichhane, S., and Fukao, T. (2021). Overlapping and stress-specific transcriptomic and hormonal responses to flooding and drought in soybean. *Plant J.* 107, 100–117. doi: 10.1111/tpj.15276

Timón-Gómez, A., Bartley-Dier, E. L., Fontanesi, F., and Barrientos, A. (2020). HIGD-driven regulation of cytochrome c oxidase biogenesis and function. *Cells* 9, 2620. doi: 10.3390/cells9122620

Timón-Gómez, A., Scharr, A. L., Wong, N. Y., Ni, E., Roy, A., Liu, M., et al. (2022). Tissue-specific mitochondrial HIGD1C promotes oxygen sensitivity in carotid body chemoreceptors. *Elife* 11, e78915. doi: 10.7554/eLife.78915.sa2

Turkan, I. (2018). ROS and RNS: key signalling molecules in plants. *J Exp Bot.* 69 (14), 3313–3315. doi: 10.1093/jxb/ery198

Valliyodan, B., Ye, H., Song, L., Murphy, M., Shannon, J. G., and Nguyen, H. T. (2017). Genetic diversity and genomic strategies for improving drought and waterlogging tolerance in soybeans. *J. Exp. Bot.* 68, 1835–1849. doi: 10.1093/jxb/erw433

Wan, S., Mak, M.-W., and Kung, S.-Y. (2012). mGOASVM: Multi-label protein subcellular localization based on gene ontology and support vector machines. *BMC Bioinf.* 13, 1–16. doi: 10.1186/1471-2105-13-290

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. doi: 10.1093/nar/gkr1293

Ye, H., Song, L., Chen, H., Valliyodan, B., Cheng, P., Ali, L., et al. (2018). A major natural genetic variation associated with root system architecture and plasticity improves waterlogging tolerance and yield in soybean. *Plant Cell Environ.* 41, 2169–2182. doi: 10.1111/pce.13190

Yu, C. S., Chen, Y. C., Lu, C. H., and Hwang, J. K. (2006). Prediction of protein subcellular localization. *Proteins: Struct. Funct. Bioinf.* 64, 643–651. doi: 10.1002/prot.21018

Yu, J., Zhang, Y., Zhong, J., Ding, H., Zheng, X., Wang, Z., et al. (2020). Water-level alterations modified nitrogen cycling across sediment-water interface in the Three Gorges Reservoir. *Environ. Sci. Pollut. Res.* 27, 25886–25898. doi: 10.1007/s11356-019-06656-z

Zhou, J., Mou, H., Zhou, J., Ali, M. L., Ye, H., Chen, P., et al. (2021). Qualification of soybean responses to flooding stress using UAV-based imagery and deep learning. *Plant Phenom.* 2021, 9892570. doi: 10.34133/2021/9892570

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 08 July 2024
DOI 10.3389/fpls.2024.1437107

OPEN ACCESS

EDITED BY
Huatao Chen,
Jiangsu Academy of Agricultural Sciences (JAAS), China

REVIEWED BY
Hengyou Zhang,
Chinese Academy of Sciences (CAS), China
Chengfu Su,
Qingdao Agricultural University, China

*CORRESPONDENCE
Qingshan Chen
qshchen@126.com
Shuming Wang
shumingw@263.net
Hongwei Jiang
j3994102@126.com

†These authors have contributed equally to this work and share first authorship

RECEIVED 23 May 2024
ACCEPTED 25 June 2024
PUBLISHED 08 July 2024

CITATION
Li G, Xie J, Zhang W, Meng F, Yang M, Fan X, Sun X, Zheng Y, Zhang Y, Wang M, Chen Q, Wang S and Jiang H (2024) Integrated examination of the transcriptome and metabolome of the gene expression response and metabolite accumulation in soybean seeds for seed storability under aging stress. *Front. Plant Sci.* 15:1437107. doi: 10.3389/fpls.2024.1437107

Integrated examination of the transcriptome and metabolome of the gene expression response and metabolite accumulation in soybean seeds for seed storability under aging stress

Guang Li[1†], Jianguo Xie[1†], Wei Zhang[1], Fanfan Meng[1], Mingliang Yang[2], Xuhong Fan[1], Xingmiao Sun[1], Yuhong Zheng[1], Yunfeng Zhang[1], Mingliang Wang[1], Qingshan Chen[2*], Shuming Wang[1*] and Hongwei Jiang[1*]

[1]Jilin Academy of Agricultural Sciences (China Agricultural Science and Technology Northeast Innovation Center), Soybean Research Institute, Changchun, China, [2]Northeast Agricultural University, Harbin, Heilongjiang, China

Soybean quality and production are determined by seed viability. A seed's capacity to sustain germination via dry storage is known as its seed life. Thus, one of the main objectives for breeders is to preserve genetic variety and gather germplasm resources. However, seed quality and germplasm preservation have become significant obstacles. In this study, four artificially simulated aging treatment groups were set for 0, 24, 72, and 120 hours. Following an aging stress treatment, the transcriptome and metabolome data were compared in two soybean lines with notable differences in seed vigor—R31 (aging sensitive) and R80 (aging tolerant). The results showed that 83 (38 upregulated and 45 downregulated), 30 (19 upregulated and 11 downregulated), 90 (52 upregulated and 38 downregulated), and 54 (25 upregulated and 29 downregulated) DEGs were differentially expressed, respectively. A total of 62 (29 upregulated and 33 downregulated), 94 (49 upregulated and 45 downregulated), 91 (53 upregulated and 38 downregulated), and 135 (111 upregulated and 24 downregulated) differential metabolites accumulated. Combining the results of transcriptome and metabolome investigations demonstrated that the difference between R31 and R80 responses to aging stress was caused by genes related to phenylpropanoid metabolism pathway, which is linked to the seed metabolite caffeic acid. According to this study's preliminary findings, the aging-resistant line accumulated more caffeic acid than the aging-sensitive line, which improved its capacity to block lipoxygenase (LOX) activity. An enzyme activity inhibition test was used to demonstrate the effect of caffeic acid. After soaking seeds in 1 mM caffeic acid (a LOX inhibitor) for 6 hours and artificially aging them for 24 hours, the germination rates of the R31 and R80

seeds were enhanced. In conclusion, caffeic acid has been shown to partially mitigate the negative effects of soybean seed aging stress and to improve seed vitality. This finding should serve as a theoretical foundation for future research on the aging mechanism of soybean seeds.

KEYWORDS

soybean, aging, caffeic acid, transcriptome, metabolomic

1 Introduction

Soybean (*Glycine max* L. Merr.), the most important agricultural legume, was first planted in China around 5000 years ago (Sedivy et al., 2017). Today, it is grown all over the world and supplies 28% of vegetable oil and 70% of the protein meal consumed globally (SoyStats, 2021). Since soybean has such high nutritional and economic importance, conserving its genetic variety and gathering its germplasm resources have been the top goals for breeders (Lin et al., 2022). To address the rising need for plant proteins, oils, and food, however, we must breed soybean germplasm with improved performance due to climate change and population expansion.

An important component of sustainable agricultural production is the viability and longevity of high-quality seeds during storage, and seed longevity is the ability of seeds to germinate after they have been stored dry. The ripening and storage of seeds is a complicated process affected by several internal and external influences (Ramtekey et al., 2022). The seed is the primary means of plant reproduction and represents a crucial developmental stage with several unique characteristics. The preservation of plant biodiversity and the success of crops are both significantly hampered by seed life. Seeds have a variety of mechanisms (protection, detoxification, and repair) to survive in dry conditions and maintain a high germination capacity. As a result, the seed system offers a useful model for researching lifespan and aging (Rajjou and Debeaujon, 2008).

The vigor of crop seeds is crucial for maintaining the germplasm and enhancing grain quality. Wang et al. conducted a thorough analysis of the transcriptome and metabonomics of two subspecies of rice with varying levels of seed vigor obtained through sped-up senescence. They discovered that *bZIP23* is most likely to influence seed vigor through a common pathway with *PER1A* and that overexpressing and knocking out these two genes increased and decreased seed vigor, respectively (Wang et al., 2022b). For crop yield, resistance to seed aging and quick seedling development are crucial agronomic features. In comparison to the null segregant (NS) control, maize seedlings grew more quickly after germination due to the hyperaccumulation of IAA in the zygotic embryo of *zmdreb2a*. Additionally, the *zmdreb2a* seeds showed reduced seed aging tolerance due to reduced raffinose levels and decreased expression of *RAFFINOSE SYNTHASE (ZmRAFS)* in their embryos (Han et al., 2020). The longevity and vigor of seeds during seed maturation and germination in peas, soybeans, and *Medicago truncatula* are determined by *RFO* levels and the expression of genes that influence its synthesis, such as ABI5, raffinose synthase, and galactinol synthase (Salvi et al., 2016; Zinsmeister et al., 2016; Pereira Lima et al., 2017).

Numerous difficulties with manufacturing, post-harvest storage, and subsequent quality are always present in seeds. In addition, due to climate change, various stressors may result in subpar seed performance, such as decreased germination, uneven seedling emergence, subpar seedling establishment, and destructive changes in the root cell structure, significantly reducing yield (Reed et al., 2022).

Seeds may age more slowly in the wild than in artificial environments, which makes it difficult to understand the physiological mechanism. Thus, the normal aging process of mimicked seeds is accelerated by artificial conditions (Wang et al., 2016, 2022). It offers a scientific foundation for an in-depth examination of seed physiology and quality control procedures, and it aids in our understanding of the physiological changes that occur in seeds as they mature (Ku et al., 2014).

An increased storage time reduces seed vigor, and high-vigor seeds have better yield potential than low-vigor seeds. Utilizing artificial aging technology, we examined the physiological alterations and associated molecular processes of soybean seed aging by artificially aging two soybean lines with varying levels of vigor ('R80' and 'R31'). We also combined transcriptome and metabolomics data analysis. This work offers fresh perspectives on safeguarding and utilizing germplasm resources and serves as a theoretical foundation for further research on the biology of soybean seeds.

2 Materials and methods

2.1 Plant materials and artificial aging conditions

The wild soybean ZYD00006 was used as the recipient parent, and the soybean cultivar SN14 as the recurrent parent. A population of 213 chromosome segment substitution lines (CSSL) was assembled using hybridization, backcrossing, and selfing. In 2020,

the population was seeded in an experimental field in Gongzhuling City, Jilin Province, China. The following were used: a randomized block design, 15 cm plant spacing, 65 cm row spacing, and field management based on traditional soybean production in the area. To mimic the aging process, an artificial aging box was employed. A humidifier holding 4–5 L of distilled or filtered water was connected to a thermostat. From each variety, 600 seeds of a full and uniform size were randomly chosen, packaged in small nylon mesh bags, and placed on the net rack in an artificial seed aging box (LH-150S) (Xin et al., 2016; Zheng et al., 2022).

Before use, the aging box was cleaned with 75% ethanol. The appropriate quantity of sterilized or purified water was added to the water tank, and the age box was opened. The temperature and humidity for the aging period were set at 45°C and 95% relative humidity, respectively. The aging process occurred for 24, 72, and 120 hours. Finally, the seeds were air-dried naturally.

A germination test was conducted and significantly changed based on the germination conditions in the "International Seed Inspection Regulations." A solution of 4% sodium hypochlorite was used to disinfect the seeds for 20–30 seconds before rinsing them 3 times in sterilized water. Forty soybean seeds from each variety were randomly chosen and placed in a glass Petri dish. Water was slowly added until a thin water film was visible on the paper. The seeds were covered with a layer of filter paper, and more water was added to moisten it (three instances). In an artificial incubator kept at a constant temperature of 25°C and in complete darkness, seeds were allowed to germinate. The number of germinated seeds was counted every day. The water absorbed by the seeds was replaced with sterile water. Germination was determined as follows: the radical be longer than half of the seed; if the radicle was spiral, the seed was not counted as germinated; and the number of germinated seeds was recorded each day for 7 days. Based on the results of standard germination tests and germination tests conducted after 96 hours of artificial accelerated aging (Supplementary Table S1), a total of 213 CSSL soybean populations were used as experimental materials. The stable phenotypic soybean lines R80 and R31, which had the highest anti-aging and aging sensitivity, respectively, were screened. Transcriptomic and metabolomic analyses of unaged R80 and R31 seeds and R31 and R80 seeds aged 24, 72, and 120 hours were performed. For each group, three biological repetitions were conducted for transcriptomic, and six biological repetitions were conducted for metabolomic.

2.2 RNA isolation and sequence analysis

Following the manufacturer's instructions, total RNA was extracted using TRIzol reagent (Invitrogen, Carlsbad, CA, USA). Using Nanodrop, Qubit 2.0, and Agilent 2100 devices, the purity, concentration, and integrity of the RNA samples, respectively, were determined. The poly(A) technique was employed to enhance the mRNA once the samples were qualified. After the mRNA was reverse transcribed using oligo (dT) primers, the cDNA was broken apart. RNA-seq was performed by Biomarker Bioinformatics Technology (China) using a Hiseq 4000 PE150 sequencing technology to sequence and analyze the RNA samples (Illumina, San Diego, CA, USA). HISAT2 (Kim et al., 2015) (version 2.0.5) with default settings was used to map the raw sequencing reads to the soybean genome (v2) after being filtered using FASTP (Chen et al., 2018). DESeq2 (version 1.20.0) (Love et al., 2014) was used to identify differentially expressed genes (DEGs) (defined as those with an absolute value of expression fold change ≥ 2 and an FDR ≤ 0.05). Supplementary Table S2 provides the primers used for qPCR to concurrently measure the transcript levels of genes associated with lipoxygenase (LOX).

2.3 Metabolite profiling

Biomarker Biotechnology Co., Ltd. (Beijing, China) conducted the non-targeted metabolome study. In summary, 300 µL of 75% methanol/water was added to 100 mg of material in a 1.5 mL centrifuge tube, and the mixture was centrifuged at 12,000 rpm for 10 minutes at 4°C. The Metlin database was used to identify every metabolite. An orthogonal partial least squares-discriminant analysis model was used to identify the differential metabolites. It had a variable importance of projection (VIP) score of ≥ 1 and a |log2 (fold change)| of ≥1. Utilizing the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (http://www.kegg.jp/kegg/compound/), the functional annotations of these metabolites were acquired.

2.4 Enzyme, metabolite, and gene expression changes as seeds age

H_2O_2 content and peroxidase (POD) activity were first examined to evaluate aging-related cell damage or seed degeneration. Seeds (0.2 g) were used to determine the H_2O_2 concentration following the technique of Doulis et al (Doulis et al., 1997), and the results were computed as mol H_2O_2 decomposition min/1 g/1 FW. The thiobarbituric acid reaction technique, as reported by Gao et al. (Gao et al., 2008), was used to measure POD activity, which was determined using assay kits (Comin, Suzhou, China). LOX activities contributed to seed healing mechanisms. To obtain enzyme crude extract with 0.1 mM potassium phosphate buffer, three replications of the same procedure were performed using approximately 0.1 g of seed each time (pH 7.8). When performing enzyme activity tests, the supernatant was kept at 4°C. As previously mentioned, sodium phosphate reaction buffer (150 mM, pH 8.0) and linoleic acid substrate solution (10 mM linoleic acid) were produced for the LOX tests (Stephany et al., 2015). A UV spectrophotometer was used to measure the LOX reaction of 0.1 g of seeds at 234 nm.

3 Results

3.1 Dissecting the transcriptome profiles of seed with different levels of vigor

This study's artificial accelerated aging induction experiment was used to assess the seed vigor. According to the characteristics of

seed germination on aging stress, R31 lost its seed germination ability, while R80 maintained a cumulative germination rate of 64.44% after 96 hours of treatment (Table 1). Based on the findings, R31 was an aging-sensitive line, and R80 was an aging-resistant line.

The variation in seed vigor between R31 and R80 may be influenced by changes in gene expression. The correlation study revealed that at various stages of artificial aging stress, all biological repetitions of the group internal sample of the R80 or R31 lines showed relatively consistent gene expression levels (Supplementary Table S3). Our study examined the expression profiles of two soybean lines (R31 and R80) that differed in their ability to withstand the effects of aging. Four comparison groups were created: R31 0 h versus R80 0 h, R31 24 h versus R80 24 h, R31 72 h versus R80 72 h, and R31 120 h versus R80 120 h. In the four comparison groups, 83 (38 upregulated and 45 downregulated), 30 (19 upregulated and 11 downregulated), 90 (52 upregulated and 38 downregulated), and 54 (25 upregulated and 29 downregulated) differentially expressed genes (DEGs) were identified respectively (Figure 1A; Supplementary Table S4).

The four artificial aging therapy groups shared 13 DEGs (Figure 1B). In addition, 12 of the 13 DEGs consistently exhibited upregulation and just one consistently showed downregulation in each comparison group (Supplementary Figure S1; Supplementary Table S4). KEGG functional enrichment analysis was then performed on the DEGs between R80 and R31. The DEGs between the non-aged groups (R80-0 h versus R31-0 h) were mainly enriched in pathways including flavonoid biosynthesis, circadian rhythm—plant, RNA polymerase, pantothenate and CoA biosynthesis, phosphonate and phosphinate metabolism, limonene and pinene degradation, glycosphingolipid biosynthesis—globo and isoglobo series, flavone and flavonol biosynthesis, histidine metabolism, sulfur metabolism, tryptophan metabolism, sphingolipid metabolism, isoflavonoid biosynthesis, beta-alanine metabolism, pyruvate metabolism, terpenoid backbone biosynthesis, lysine degradation, fatty acid degradation, valine, leucine and isoleucine degradation, and glycerolipid metabolism (Figure 2). DEGs between aged seed groups [(R31 0 h versus R80 0 h, R31 24 h versus R80 24 h, R31 72 h versus R80 72 h, and R31 120 h versus R80 120 h) were mostly related to phosphonate and phosphinate metabolism, MAPK signaling pathway—plant, glycerolipid metabolism, valine, leucine and isoleucine degradation, fatty acid degradation, lysine degradation, pyruvate metabolism, beta-alanine metabolism, sphingolipid metabolism, sulfur metabolism, histidine metabolism, glycosphingolipid biosynthesis—globo and isoglobo series, limonene and pinene degradation, pantothenate and CoA biosynthesis, RNA polymerase, circadian rhythm—plant, plant hormone signal transduction, spliceosome, protein processing in endoplasmic reticulum, pentose and glucuronate interconversions, galactose metabolism, inositol phosphate metabolism, glycosylphosphatidylinositol—anchor biosynthesis, and ascorbate and aldarate metabolism (Figure 2). Remarkably, the KEGG enrichment study (Figure 2), before and after artificial aging therapy, showed that R31 and R80 had different metabolic pathways, especially with regard to flavonoid, flavonoid, and flavanol metabolism. Between the two lines, these metabolic pathways continuously varied (Table 2; Figure 2; Supplementary Figure S2). The transcriptome study also revealed that aging treatment caused a greater transcriptional difference between R80 and R31 seeds, explaining the increase in anti-aging processes in R80 seeds.

3.2 KEGG analysis of differentially accumulated metabolites in seed during aging stress

The purpose of artificially accelerating aging was to identify the metabolite differences between the R31 and R80 seeds. The R31 and R80 seeds used in this study were matured for 0, 24, 72, and 120 hours before being used as samples for Qualcomm quantitative metabolite analysis. After metabolomics data analysis, 799 metabolites were discovered. The correlation study revealed that at different stages of artificial aging stress, all biological repetitions of the group internal sample of the R80 or R31 lines showed relatively consistent accumulated metabolite patterns (Supplementary Table S5). Four comparison groups were created and measured in total. The control group was the non-aged group, and the treatment groups comprised three artificially aged groups, with aging for 24, 72, and 120 hours. There were 62 (29 upregulated and 33 downregulated), 94 (49 upregulated and 45 downregulated), 91 (53 upregulated and 38 downregulated), and 135 (111 upregulated and 24 downregulated) differentially accumulated metabolites (DAMs) in the four comparison groups, respectively, identified using a VIP score of 1 and an absolute multiple change of 2 (Figure 3).

The KEGG enrichment indicates that the aging process of the R31 and R80 seeds drastically changed several metabolic pathways. These pathways included phenylpropanoid biosynthesis, flavonoid biosynthesis, tryptophan metabolism, diterpenoid biosynthesis,

TABLE 1 Modification of seed germination in R31 and R80 after accelerated aging treatment.

Treatment	ID	Repetition1	Repetition2	Repetition3	Average germination rate
CK	R31	90.00%	93.33%	100.00%	94.44%
CK	R80	93.33%	96.67%	70.00%	86.67%
Aging	R31	0.00%	0.00%	0.00%	0.00%
Aging	R80	60.00%	63.33%	70.00%	64.44%

The seeds were aged at 95% RH and below 45°C.

10.3389/fpls.2024.1437107

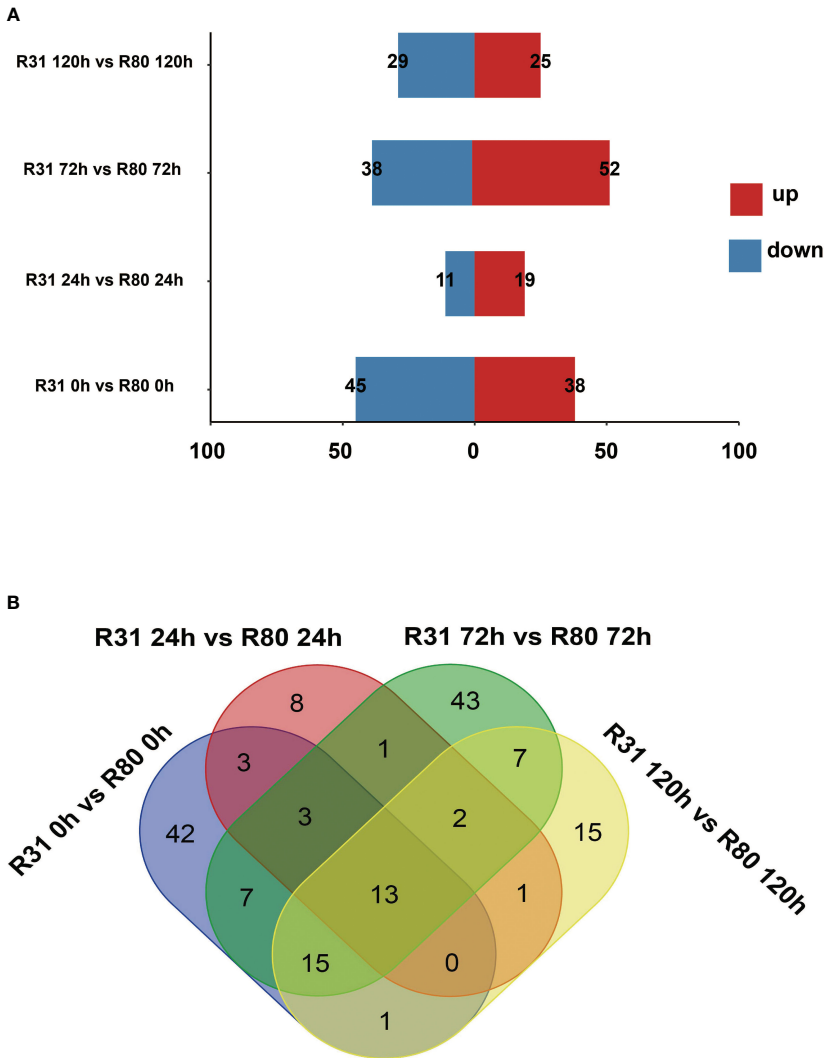


FIGURE 1
Statistics on the number of differentially expressed genes (DEGs) in seeds under aging stress **(A)**; Venn diagram of changes in differentially expressed genes (DEGs) in R31 and R80 seeds under aging stress **(B)**.

fatty acid degradation, fatty acid elongation, linoleic acid metabolism, biosynthesis of unsaturated fatty acids, alanine, aspartate, and glutamate metabolism (Figure 4). With the use of metabolic analysis network technologies, a metabolic network was created to investigate any possible relationships between these metabolites. These metabolites are associated with the key nodes of the network involved in phenylpropanoid biosynthesis, flavonoid biosynthesis, isoflavonoid biosynthesis, flavone biosynthesis, and flavonol biosynthesis (Figure 5A). Several cumulative metabolites (DAMs) were identified, including caffeic acid, coumarin, sinapic acid, scopoletin, 4-hydroxy-3-methoxycinnamaldehyde, and isoliquiritigenin (Figures 5A; Supplementary Table S6). Unexpectedly, during aging stress, phenylpropanoid and flavonoid levels rose, particularly during dynamic variations in the concentration of caffeic acid. The results indicated that the caffeic acid concentration of the R80 seeds exhibited an increasing tendency and remained greater than that of the R31 seeds with the extension of artificial aging time (Figure 5B). The experimental data mentioned earlier suggest that caffeic acid may be a significant factor in the functional modulation of seed vigor under aging stress.

3.3 Activities of LOX and antioxidant enzymes

Reactive oxygen species (ROS) accumulation, ascorbic acid-glutathione circulatory activity decline, and mitochondrial function delay all contribute to mitochondrial dysfunction in aged soybean seeds (Xin et al., 2014). Excess ROS buildup in mitochondria results in the breakdown of the antioxidant system, which is the cause of seed degeneration (Kurek et al., 2019). Two soybean cultivars, 'R80' and 'R31,' were shown to have endogenous H_2O_2 levels that increased with age. Notably, R31 seeds did not exhibit aging resistance, but had much greater endogenous H2O2 concentrations than R80 seeds. Compared to the sensitive aging variety R31, the endogenous H2O2 level of age-resistant variety R80

10.3389/fpls.2024.1437107

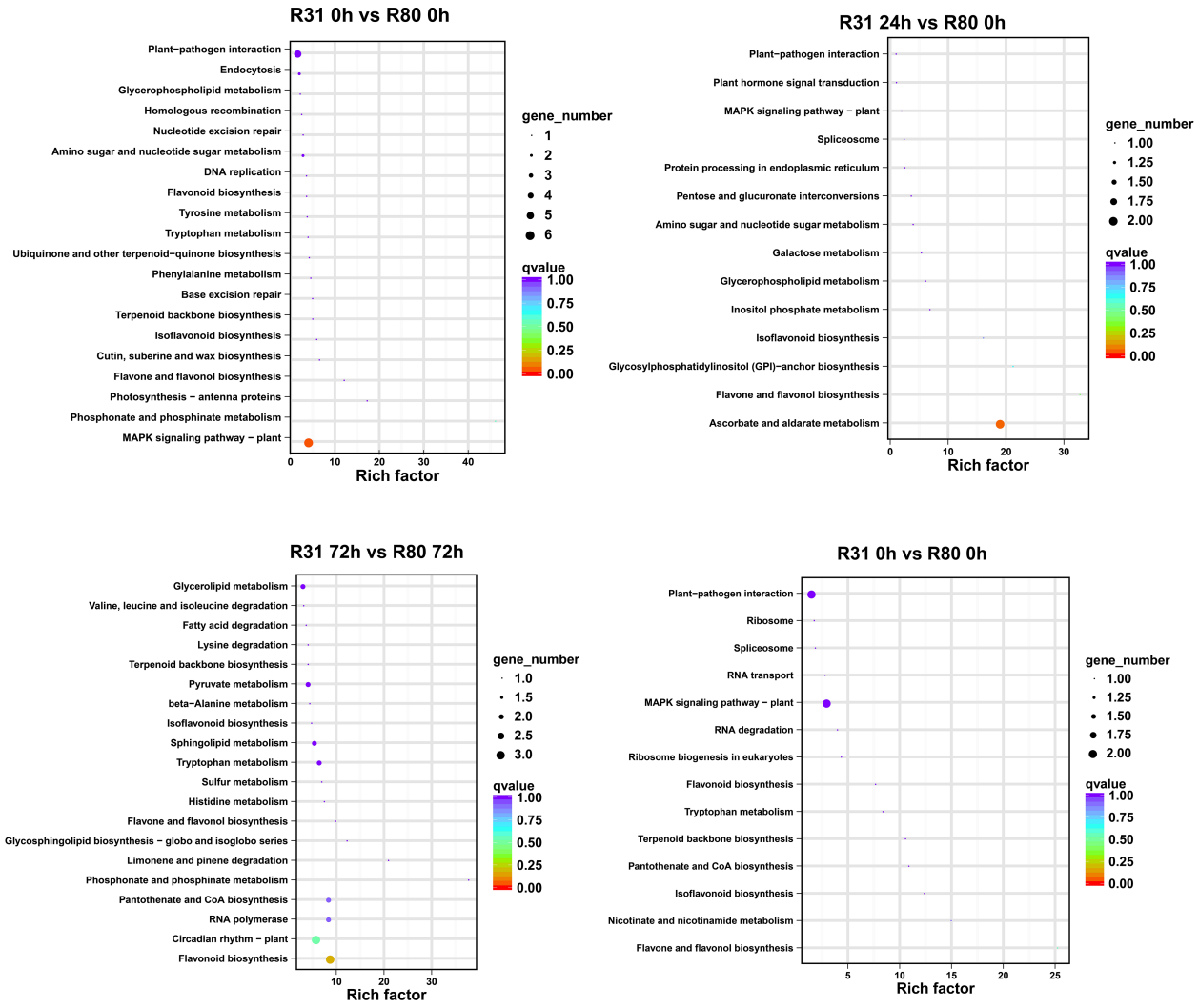


FIGURE 2
KEGG pathway changes in aged seeds of R31 and R80.

decreased by 5.28, 47.78, and 33.09% after rapid artificial aging for 24, 72, and 120 hours, respectively. In contrast, the endogenous POD level of the aging-tolerant variety R80 increased by 55.56, 19.92, and 12.93% after rapid artificial aging for 24, 72, and 120 hours, respectively, compared to the sensitive aging variety R31. Enzymes play crucial roles in the growth and development of plant life. The LOX family of enzymes is one of the most important. During seed storage, LOX can catalyze the oxidation of unsaturated fatty acids and produce hydroperoxide, which reduces the vigor and nutritional quality of seeds (Viswanath et al., 2020). Seed vigor, antioxidant enzymes, and LOX activities, along with corresponding gene expression, were evaluated after aging to gather more insight into the possible mechanisms behind seed deterioration and to validate the idea that LOX activity might be a new sensitive signal

TABLE 2 List of DEGs involved in phenylpropanoid metabolism pathway during aging of R31 and R80 seeds.

Gene_name	R31 0 h versus R80 0 h	R31 24 h versus R80 24 h	R31 72 h versus R80 72 h	R31 72 h versus R80 72 h	KEGG_pathway_annotation
Glyma.19G254600	Down	Normal	Down	Down	Flavonoid biosynthesis
Glyma.08G247100	Up	Up	Up	Up	Isoflavonoid biosynthesis; flavone and flavonol biosynthesis
Glyma.01G228700	Normal	Normal	Up	Normal	Flavonoid biosynthesis; circadian rhythm—plant
Glyma.11G011500	Normal	Normal	Up	Normal	Flavonoid biosynthesis; circadian rhythm—plant

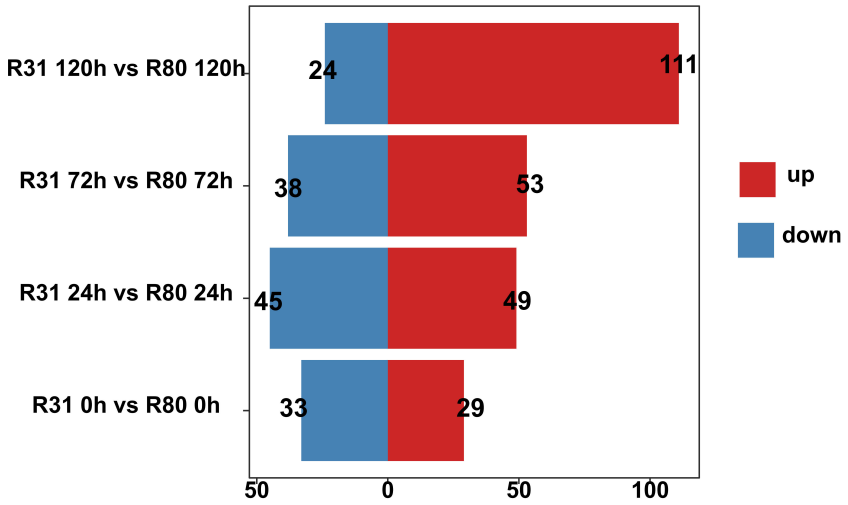


FIGURE 3
Statistics on the quantity of differentially accumulated metabolites for seeds under aging stress.

for predicting seed aging during storage. According to the findings, under artificially generated aging stress, the R80 seeds showed reduced LOX activity compared with the R31 seeds (Figure 6A). Moreover, the R80 seeds had lower levels of LOX gene expression than the R31 seeds (Figure 7A). LOX activity suppression research is required to comprehend how soybean seeds age and to find a way to postpone seed aging and extend seed life. Furthermore, LOX activity was considerably inhibited by a few phenolic substances, with caffeic acid being the most potent inhibitor (approximately 57% of inhibition) (Szymanowska et al., 2009). Additionally, by soaking R31 and R80 in 1 mM caffeic acid for 6 hours and artificially aging them for 24 hours, the germination rates of these seeds were raised. These results showed that LOX promoted the aging of soybean seeds and that reducing LOX activity preserved the vitality and viability of aging seeds.

3.4 qRT-PCR verified the accuracy of RNA-seq in detecting the gene expression level

In this work, the accuracy of transcriptome data in determining gene expression levels was confirmed using qRT-PCR. Four genes encoding DEGs and five genes encoding LOX were chosen at random. The findings demonstrated that the transcriptome data

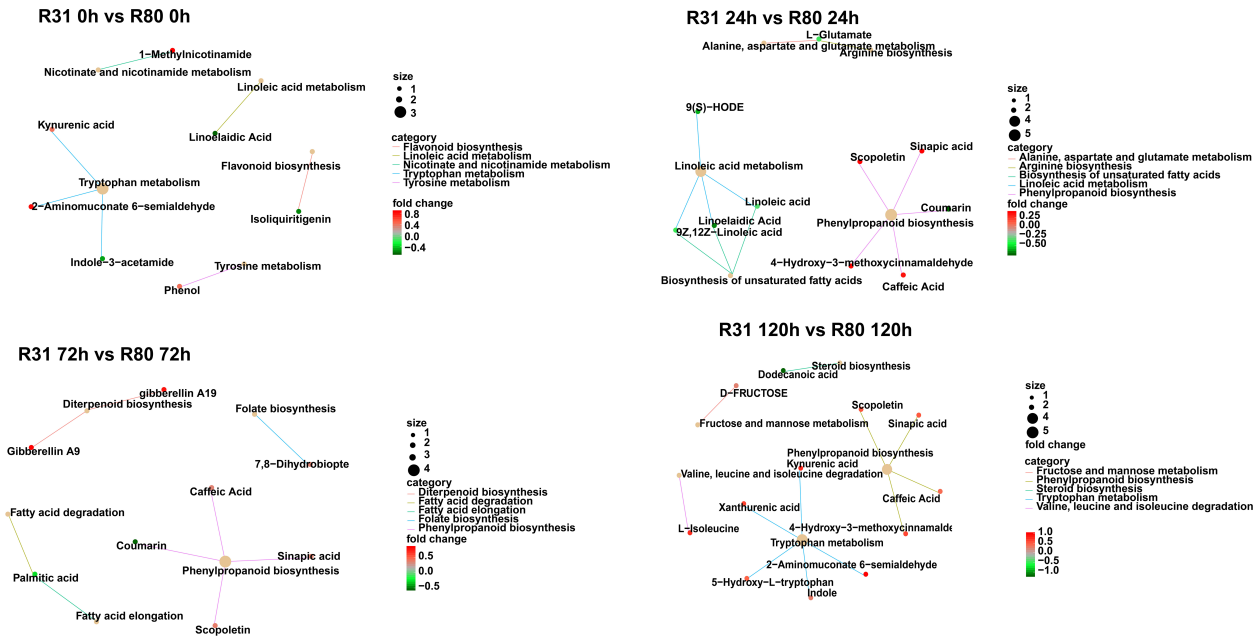


FIGURE 4
KEGG analysis of differentially accumulated metabolites (DAMs) in R31 and R80 seeds under aging stress.

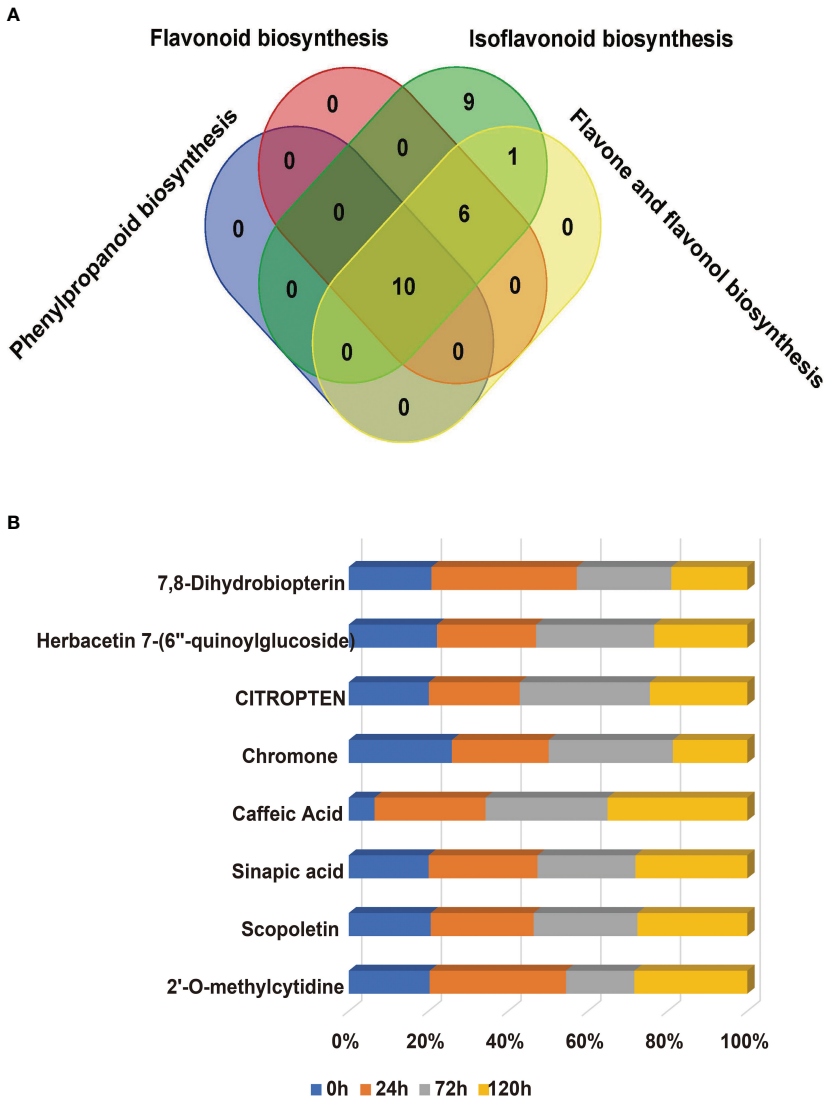


FIGURE 5
Venn diagram of the changes of differentially accumulated metabolites (DAMs) related to phenylpropanoid metabolism pathway in R31 and R80 seeds under aging stress **(A)**; Compared with R31, the increasing proportion of different accumulated metabolites in R80 changed with aging time **(B)**.

of these chosen genes were highly consistent with the qRT-PCR results (Figure 7A) and that the qRT-PCR results and the relative expression level of log $_{10}$(FPKM) of nine genes obtained by RNA-seq exhibited a nearly linear correlation (Figure 7B).

4 Discussion

Seeds are essential for crop growth, human nutrition, and food security. The primary elements influencing crop seed function are the intricate features of seed vigor. Successful planting is critical to crop productivity and resource efficiency. The strength, regularity, and speed at which seeds germinate and generate seedlings under a variety of environmental conditions are determined by their vigor (Finch-Savage and Bassel, 2015). A chemical fingerprint, or signature of an organism's or sample's metabolic condition at a particular moment in time, can be obtained through metabolomic analysis. The metabolites to be examined may be the end or intermediate products of plant metabolic pathways, or they may arise due to environmental influences from the outside world (Shen et al., 2022). Extreme flexibility in phenylpropanoid metabolism can result in a large variety of products that function in plant growth and interactions with the environment in response to various developmental stages and constantly shifting environmental variables (Yuan and Grotewold, 2020). Seed permeability and resistance to mechanical damage are correlated with the lignin concentration of the soybean seed coat (Capeleti et al., 2005).

The seed coat is a type of composite structure that may be used as a conduit for nourishment obtained from growing embryos. The seed coat offers the embryos shelter and protection once the seeds have dried and matured. It can also apply dormancy or cause germination by regulating water absorption. The qualities of the

10.3389/fpls.2024.1437107

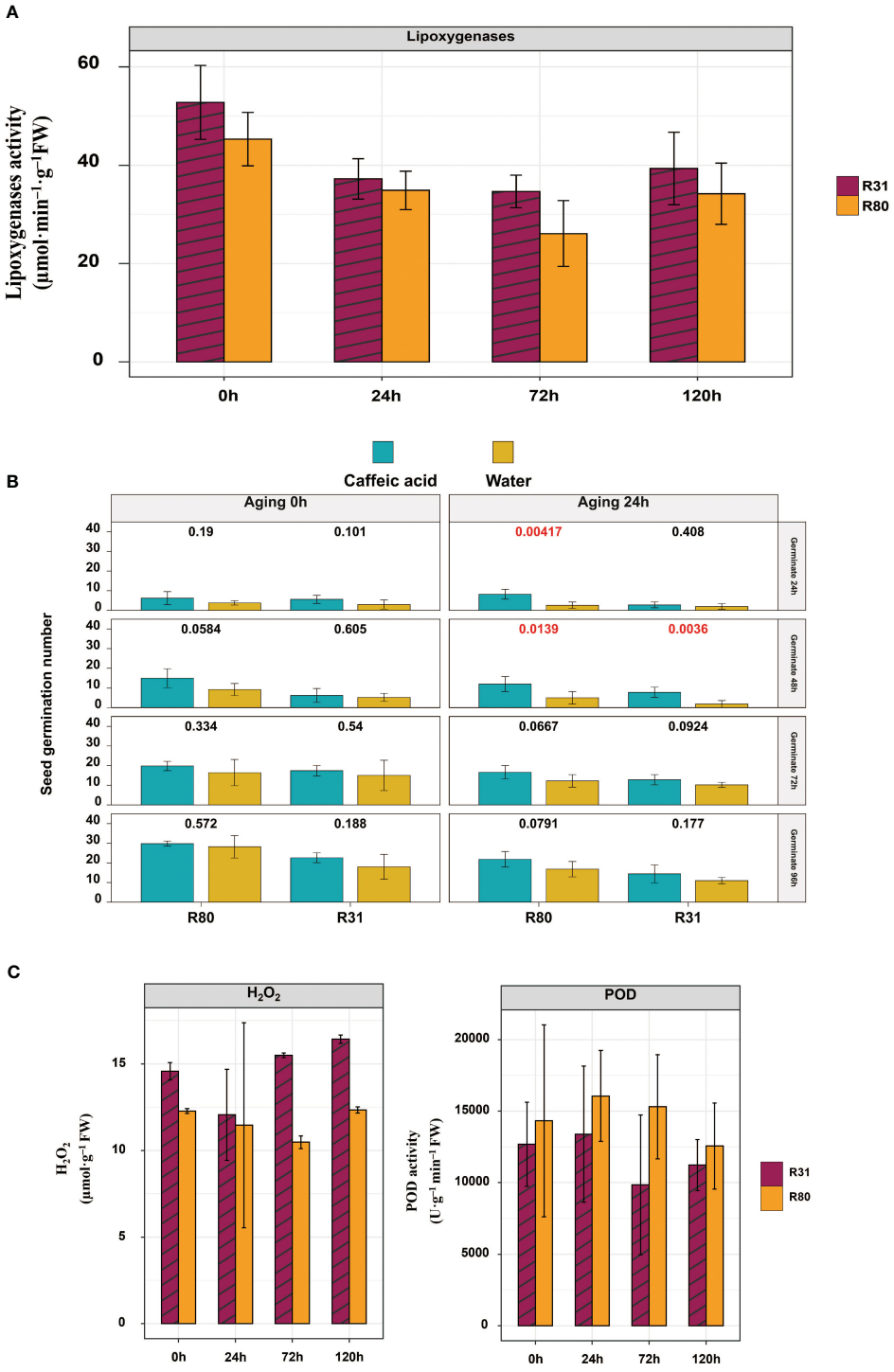


FIGURE 6
Lipoxygenase (LOX) activity of R31 and R80 seeds during aging stress **(A)**; inhibition of LOX activity increased the vigor of R31 and R80 seeds under artificial aging stress (seed number n = 30) **(B)**; and seed aging increased the contents of endogenous H_2O_2 and POD in R31 and R80 seeds **(C)**.

seeds as a whole and the usefulness of their derivatives will be influenced by the features established by the seed coat for crops such as soybean. Recently, fascinating instances of atypical genetic pathways regulating seed coat breaking, gloss, and color have been identified in soybean (Qutob et al., 2008). Plant physiology depends on a large class of secondary plant metabolites generated from phenylalanine. Seed coats contain phenolic chemicals that make the seed harder and prevent microbial development. The seed coat shields the seed from electrolyte leaks and hydration stress during germination (Mohamed-Yasseen et al., 1994). In the current investigation, we found that the seed vigor of line R80 was superior to that of line R31 under the stress of aging (Table 1). Natural phenolic compounds include phenylpropanoids. They are a broad group of phenylalanine-derived secondary plant metabolites

10.3389/fpls.2024.1437107

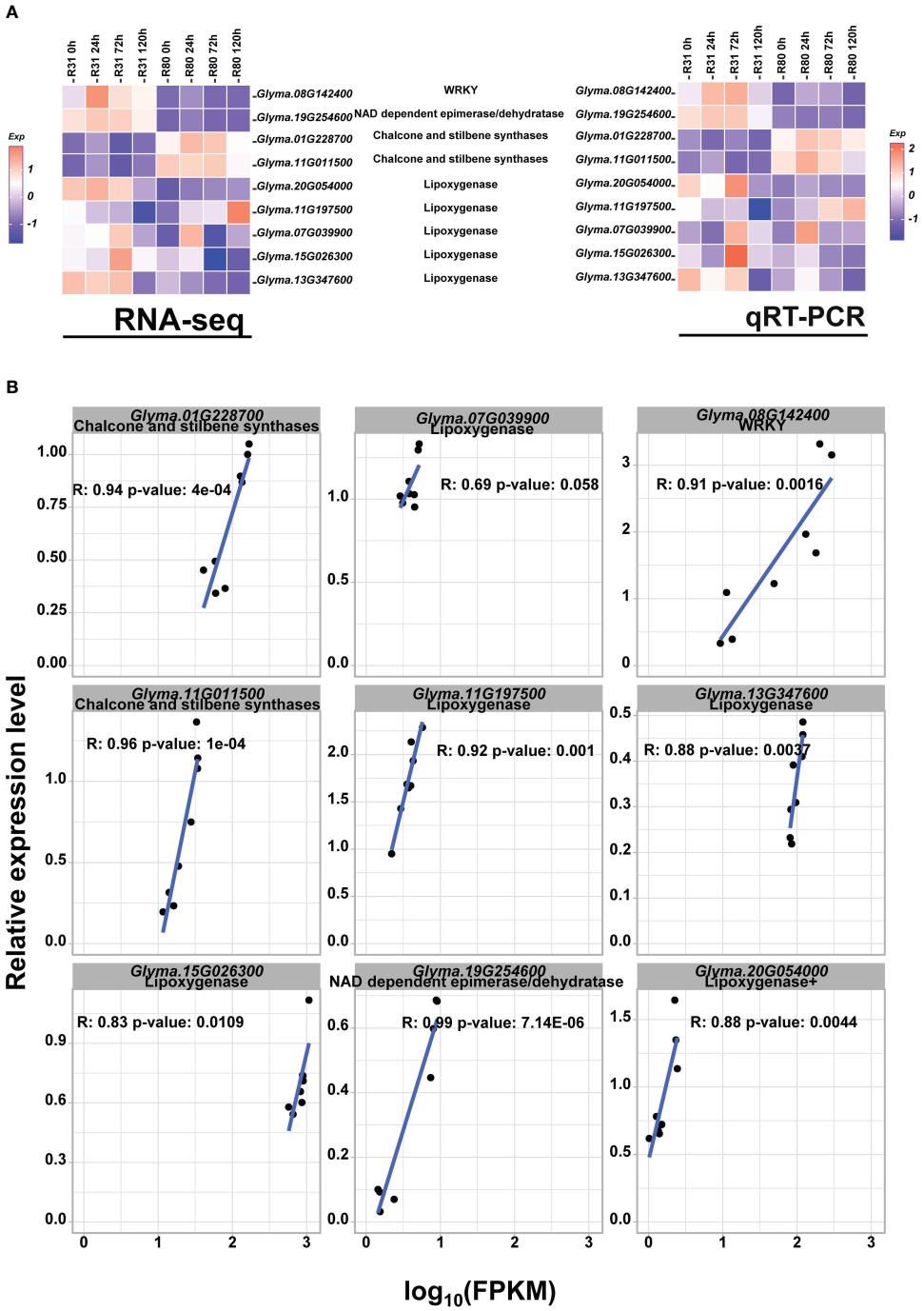


FIGURE 7
qRT-PCR verified the accuracy of RNA-seq in detecting gene expression levels. Heatmap of RNA-seq and RT-PCR gene expression levels for the randomly selected DEGs. **(A)**; Correlation analysis of RNA-seq and RT-PCR gene expression levels for the randomly selected DEGs **(B)**.

that are essential to plant physiology (Dong and Lin, 2021). These substances serve as crucial cell wall building blocks, shielding plants from a variety of biotic and abiotic environmental stressors (Treutter, 2006; Vogt, 2010; Kiani et al., 2021). Integrating the data from transcriptome and metabolome analyses showed that the DAMs and DEGs of the four comparison groups (R80-0 h versus R31-0 h, R31-24 h versus R80-24 h, R31-72 h versus R80 72 h, and R31-120 h versus R80-120 h) were mapped to the KEGG database. Phenylpropanoid, flavonoids, isoflavones, flavonoids, and flavonols were among the co-mapped pathways. Particularly those connected to the metabolite of seed coat known as caffeic acid were responsible for the difference between R31 and R80 response to aging stress (Figures 2, 4).

Seeds have evolved extraordinarily effective repair systems, including enzymatic antioxidant systems, to achieve homeostasis of H_2O_2 production (Xia et al., 2015). During lipid peroxidation in the seeds of rice (*Oryza sativa* L.) and soybean (*Glycine max* (L.) Merr.), LOX (LOX, EC1.13.11.12) plays a significant role (Lima

et al., 2010; Xu et al., 2015). When artificially generated aging stress was applied to the R80 seeds, LOX activity was lower than that in the R31 seeds (Figure 6A). Both internal metabolic processes and external stimuli cause the generation of ROS, including superoxide anion radicals, hydrogen peroxide, and hydroxyl radicals, in all cells. Nonetheless, by activating several antioxidant mechanisms, cells are often able to lower the oxidative potential of ROS. Several substances, including caffeic acid and its derivatives, have been shown to possess antioxidant qualities (Nardini et al., 2001). In artificially aged soybean seeds, LOX activity was inhibited by the application of caffeic acid. Additionally, the viability of the R80 and R31 seeds improved dramatically after 24 hours of artificial aging, suggesting that caffeic acid may strengthen the seeds' resistance to storage (Figure 6B). In contrast to the sensitive aging variety R31, the aging-tolerant variety R80 demonstrated that after rapid artificial aging for 24, 72, and 120 hours, the endogenous levels of H_2O_2 fell by 5.28, 47.78, and 33.09%, respectively (Figure 6C).

In plant research, there has been a great deal of interest in the detection and characterization of differential gene expression from tissues exposed to stress. The likelihood of advancing crop improvement by direct genetic modification increases with the identification of components involved in the response to a certain stress (Dunwell et al., 2001). Previous work discovered a novel WRKY transcription factor, *OsWRKY29*, that adversely controls rice seed dormancy. *OsWRKY29* overexpression decreased seed dormancy, whereas its knockout and RNA interference increased it (Zhou et al., 2020). For instance, transcription factors such as WRKY3 and NFLX1, which are involved in plant defense, also have an influence on seed survival by controlling the permeability of the seed coat (Debeaujon et al., 2000). This study's findings demonstrated that under normal conditions, the expression level of the WRKY (*Glyma.08G142400*) transcription factor in R80 was much lower than that of R31. Thus, we concluded that R80 exhibited low WRKY expression (Figure 7A), which helped R80 seeds become dormant and avoid the negative effects of aging stress.

As a consequence of inter-cultivar vigor fluctuations and artificial aging, we discovered numerous potential metabolites and related DEGs by the interactive comparison of transcriptomic and metabolomic data. The genetic and metabolic underpinnings of inter-cultivar vigor variations and artificial soybean seed aging are better understood because of this research.

5 Conclusion

Maintaining genetic variety and germplasm resources has become a major concern for breeders to achieve seed quality and germplasm preservation. This study compared and examined the transcription and metabolic data of two soybean lines treated with aging stress: R31 (aging sensitive) and R80 (aging tolerant). Four sets of artificial aging treatments—0, 24, 72, and 120 hours—with varying durations were performed. There were differences in the DEGs and differentially accumulated metabolites between the two soybean lines aged for different durations. The ability of soybean seeds to withstand the effects of age has been revealed to be mostly regulated by the phenylpropanoid metabolic pathway, particularly caffeic acid. Longer aging treatment periods produced higher levels of caffeic acid, and this buildup enhanced the seed's anti-aging ability by blocking LOX activity. Moreover, the germination rates of the R31 and R80 seeds were increased by immersing them in 1 mM caffeic acid for 6 hours and artificially aging them for 24 hours. Overall, this study indicates that the detrimental effects of seed aging stress may be somewhat mitigated and that seed vitality can be improved by the accumulation of metabolites, specifically caffeic acid, in the soybean phenylpropanoid acid metabolic pathway. The results of this study offer a theoretical framework for further investigations into the mechanism of soybean seed aging.

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/, PRJNA1112530.

Author contributions

GL: Writing – original draft, Visualization. JX: Writing – original draft, Data curation. WZ: Funding acquisition, Resources, Writing – review & editing. FM: Writing – review & editing, Investigation. MY: Validation, Writing – review & editing. X-HF: Formal analysis, Writing – review & editing. XS: Investigation, Writing – review & editing. YHZ: Validation, Writing – review & editing. YFZ: Resources, Writing – review & editing. MW: Resources, Writing – review & editing. QC: Funding acquisition, Writing – review & editing. SW: Funding acquisition, Writing – review & editing. HJ: Funding acquisition, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. Financial support was received from the National Science Foundation of China (U21A20215) and the China Agriculture Research System of MOF and MARA (CARS-04-PS08 and CARS-04-PS15).

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.2024.1437107/full#supplementary-material

SUPPLEMENTARY TABLE 1
Germination rate of 213 CSSL soybean populations during a standard germination test and after a 96-hour artificial accelerated aging test.

SUPPLEMENTARY TABLE 2
Gene primer sequences used in qRT-PCR.

SUPPLEMENTARY TABLE 3
Correlation matrix analyses of gene expressions among biological replicates of different samples from different artificial aging conditions between the R80 and the R31 lines.

SUPPLEMENTARY TABLE 4
Differentially expressed genes of the four comparison groups.

SUPPLEMENTARY TABLE 5
Correlation matrix analyses of accumulated metabolites among biological replicates of different samples from different artificial aging conditions between the R80 and the R31 lines.

SUPPLEMENTARY TABLE 6
Differentially accumulated metabolites of the four comparison groups.

References

Capeleti, I., Bonini, E. A., Ferrarese, M. L. L., Teixeira, A. C. N., Krzyzanowski, F. C., and Ferrarese-Filho, O. (2005). Lignin content and peroxidase activity in soybean seed coat susceptible and resistant to mechanical damage. *Acta Physiol. Plant* 27, 103–108. doi: 10.1007/s11738-005-0042-2

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics.* 34, i884–i890. doi: 10.1093/bioinformatics/bty560

Debeaujon, I., Léon-Kloosterziel, K. M., and Koornneef, M. (2000). Influence of the testa on seed dormancy, germination, and longevity in arabidopsis1. *Plant Physiol.* 122, 403–414. doi: 10.1104/pp.122.2.403

Dong, N.-Q., and Lin, H.-X. (2021). Contribution of phenylpropanoid metabolism to plant development and plant–environment interactions. *J. Integr. Plant Biol.* 63, 180–209. doi: 10.1111/jipb.13054

Doulis, A. G., Debian, N., Kingston-Smith, A. H., and Foyer, C. H. (1997). Differential localization of antioxidants in maize leaves. *Plant Physiol.* 114, 1031–1037. doi: 10.1104/pp.114.3.1031

Dunwell, J. M., Moya-León, M. A., and Herrera, R. (2001). Transcriptome analysis and crop improvement (a review). *Biol. Res.* 34, 153–164. doi: 10.4067/S0716-97602001000300003

Finch-Savage, W. E., and Bassel, G. W. (2015). Seed vigour and crop establishment: extending performance beyond adaptation. *J. Exp. Bot.* 67, 567–591. doi: 10.1093/jxb/erv490

Gao, C., Hu, J., Zhang, S., Zheng, Y., and Knapp, A. (2008). Association of polyamines in governing the chilling sensitivity of maize genotypes. *Plant Growth Regul.* 57, 31. doi: 10.1007/s10725-008-9315-2

Han, Q., Chen, K., Yan, D., Hao, G., Qi, J., Wang, C., et al. (2020). ZmDREB2A regulates ZmGH3.2 and ZmRAFS, shifting metabolism towards seed aging tolerance over seedling growth. *Plant J.* 104, 268–282. doi: 10.1111/tpj.14922

Kiani, R., Arzani, A., and Mirmohammady Maibody, S. A. M. (2021). Polyphenols, flavonoids, and antioxidant activity involved in salt tolerance in wheat, aegilops cylindrica and their amphidiploids. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.646221

Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 12, 357–360. doi: 10.1038/nmeth.3317

Ku, L., Cui, X., Cheng, F., Guo, S., Qi, J., Tian, Z., et al. (2014). Genetic dissection of seed vigour under artificial ageing conditions using two joined maize recombinant inbred line populations. *Plant Breed.* 133, 728–737. doi: 10.1111/pbr.12221

Kurek, K., Plitta-Michalak, B., and Ratajczak, E. (2019). Reactive oxygen species as potential drivers of the seed aging process. *Plants.* 8, 174. doi: 10.3390/plants8060174

Lima, W., Borem, A., Dias, D., Moreira, M., and Dias, L. (2010). Lipoxygenase and physiological quality of soybean seeds during storage. *Seed Sci. Technol.* 38, 767–771. doi: 10.15258/sst.2010.38.3.23

Lin, Y.-X., Xu, H.-J., Yin, G.-K., Zhou, Y.-C., Lu, X.-X., and Xin, X. (2022). Dynamic changes in membrane lipid metabolism and antioxidant defense during soybean (*Glycine max* L. Merr.) seed aging. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.908949

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

Mohamed-Yasseen, Y., Barringer, S. A., Splittstoesser, W. E., and Costanza, S. (1994). The role of seed coats in seed viability. *Bot. Rev.* 60, 426–439. doi: 10.1007/BF02857926

Nardini, M., Leonardi, F., Scaccini, C., and Virgili, F. (2001). Modulation of ceramide-induced NF-κB binding activity and apoptotic response by caffeic acid in U937 cells: comparison with other antioxidants. *Free Radical Biol. Med.* 30, 722–733. doi: 10.1016/S0891-5849(00)00515-3

Pereira Lima, J. J., Buitink, J., Lalanne, D., Rossi, R. F., Pelletier, S., da Silva, E. A. A., et al. (2017). Molecular characterization of the acquisition of longevity during seed maturation in soybean. *PloS One* 12, e0180282. doi: 10.1371/journal.pone.0180282

Qutob, D., Ma, F., Peterson, C. A., Bernards, M. A., and Gijzen, M. (2008). Structural and permeability properties of the soybean seed coat. *Botany.* 86, 219–227. doi: 10.1139/B08-002

Rajjou, L., and Debeaujon, I. (2008). Seed longevity: Survival and maintenance of high germination ability of dry seeds. *C. R. Biol.* 331, 796–805. doi: 10.1016/j.crvi.2008.07.021

Ramtekey, V., Cherukuri, S., Kumar, S., V, S. K., Sheoran, S., K, U. B., et al. (2022). Seed longevity in legumes: deeper insights into mechanisms and molecular perspectives. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.918206

Reed, R. C., Bradford, K. J., and Khanday, I. (2022). Seed germination and vigor: ensuring crop sustainability in a changing climate. *Heredity.* 128, 450–459. doi: 10.1038/s41437-022-00497-2

Salvi, P., Saxena, S. C., Petla, B. P., Kamble, N. U., Kaur, H., Verma, P., et al. (2016). Differentially expressed galactinol synthase(s) in chickpea are implicated in seedvigor and longevity by limiting the age induced ROS accumulation. *Sci. Rep.* 6, 35088. doi: 10.1038/srep35088

Sedivy, E. J., Wu, F., and Hanzawa, Y. (2017). Soybean domestication: the origin, genetic architecture and molecular bases. *New Phytol.* 214, 539–553. doi: 10.1111/nph.14418

Shen, S., Zhan, C., Yang, C., Fernie, A. R., and Luo, J. (2022). Metabolomics-centered mining of plant metabolic diversity and function: past decade and future perspectives. *Mol. Plant.* 16 (1), 43–63. doi: 10.1016/j.molp.2022.09.007

SoyStats (2021). American soybean associatio. Available online at: https://soygrowerscom/wp-content/uploads/2021/07/2021-Soy-Stats-WEBpdf.

Stephany, M., Bader-Mittermaier, S., Schweiggert-Weisz, U., and Carle, R. (2015). Lipoxygenase activity in different species of sweet lupin (*Lupinus* L.) seeds and flakes. *Food Chem.* 174, 400–406. doi: 10.1016/j.foodchem.2014.11.029

Szymanowska, U., Jakubczyk, A., Baraniak, B., and Kur, A. (2009). Characterisation of lipoxygenase from pea seeds (*Pisum sativum* var. *Telephone* L.). *Food Chem.* 116, 906–910. doi: 10.1016/j.foodchem.2009.03.045

Treutter, D. (2006). Significance of flavonoids in plant resistance: a review. *Environ. Chem. Lett.* 4, 147–157. doi: 10.1007/s10311-006-0068-8

Viswanath, K. K., Varakumar, P., Pamuru, R. R., Basha, S. J., Mehta, S., and Rao, A. D. (2020). Plant lipoxygenases and their role in plant physiology. *J. Plant Biol.* 63, 83–95. doi: 10.1007/s12374-020-09241-x

Vogt, T. (2010). Phenylpropanoid biosynthesis. *Mol. Plant* 3, 2–20. doi: 10.1093/mp/ssp106

Wang, B., Wang, S., Tang, Y., Jiang, L., He, W., Lin, Q., et al. (2022a). Transcriptome-wide characterization of seed aging in rice: identification of specific long-lived mRNAs for seed longevity. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.857390

Wang, B., Zhang, Z., Fu, Z., Liu, Z., Hu, Y., and Tang, J. (2016). Comparative QTL analysis of maize seed artificial aging between an immortalized F_2 population and its corresponding RILs. *Crop J.* 4, 30–39. doi: 10.1016/j.cj.2015.07.004

Wang, W.-Q., Xu, D.-Y., Sui, Y.-P., Ding, X.-H., and Song, X.-J. (2022b). A multiomic study uncovers a bZIP23-PER1A–mediated detoxification pathway to enhance seed vigor in rice. *Proc. Natl. Acad. Sci. U.S.A.* 119, e2026355119. doi: 10.1073/pnas.2026355119

Xia, F., Wang, X., Li, M., and Mao, P. (2015). Mitochondrial structural and antioxidant system responses to aging in oat (*Avena sativa* L.) seeds with different moisture contents. *Plant Physiol. Biochem.* 94, 122–129. doi: 10.1016/j.plaphy.2015.06.002

Xin, D., Qi, Z., Jiang, H., Hu, Z., Zhu, R., Hu, J., et al. (2016). and epistatic effect analysis of 100-seed weight using wild soybean (*Glycine soja* sieb. & Zucc.) chromosome segment substitution lines. *PloS One* 11, e0149380. doi: 10.1371/journal.pone.0149380

Xin, X., Tian, Q., Yin, G., Chen, X., Zhang, J., Ng, S., et al. (2014). Reduced mitochondrial and ascorbate–glutathione activity after artificial ageing in soybean seed. *J. Plant Physiol.* 171, 140–147. doi: 10.1016/j.jplph.2013.09.016

Xu, H., Wei, Y., Zhu, Y., Lian, L., Xie, H., Cai, Q., et al. (2015). Antisense suppression of LOX3 gene expression in rice endosperm enhances seed longevity. *Plant Biotechnol. J.* 13, 526–539. doi: 10.1111/pbi.12277

Yuan, L., and Grotewold, E. (2020). Plant specialized metabolism. *Plant Sci.* 298, 110579. doi: 10.1016/j.plantsci.2020.110579

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, C., Lin, Q., Lan, J., Zhang, T., Liu, X., Miao, R., et al. (2020). WRKY transcription factor osWRKY29 represses seed dormancy in rice by weakening abscisic acid response. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.00691

Zinsmeister, J., Lalanne, D., Terrasson, E., Chatelain, E., Vandecasteele, C., Vu, B. L., et al. (2016). ABI5 is a regulator of seed maturation and longevity in legumes. *Plant Cell.* 28, 2735–2754. doi: 10.1105/tpc.16.00470

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 11 July 2024
DOI 10.3389/fpls.2024.1446591

OPEN ACCESS

EDITED BY
Huatao Chen,
Jiangsu Academy of Agricultural Sciences (JAAS), China

REVIEWED BY
Dawei Xue,
Hangzhou Normal University, China
Wenxian Liu,
Lanzhou University, China

*CORRESPONDENCE
Mingjia Chen
mjchen@njau.edu.cn

†These authors have contributed equally to this work

RECEIVED 10 June 2024
ACCEPTED 25 June 2024
PUBLISHED 11 July 2024

CITATION
Liu P, Liu H, Zhao J, Yang T, Guo S, Chang L, Xiao T, Xu A, Liu X, Zhu C, Gan L and Chen M (2024) Genome-wide identification and functional analysis of mRNA m^{6}A writers in soybean under abiotic stress.
Front. Plant Sci. 15:1446591.
doi: 10.3389/fpls.2024.1446591

Genome-wide identification and functional analysis of mRNA m^{6}A writers in soybean under abiotic stress

Peng Liu[1†], Huijie Liu[1†], Jie Zhao[1†], Tengfeng Yang[1], Sichao Guo[1], Luo Chang[1], Tianyun Xiao[1], Anjie Xu[1], Xiaoye Liu[2], Changhua Zhu[1], Lijun Gan[1] and Mingjia Chen[1*]

[1]College of Life Sciences, Nanjing Agricultural University, Nanjing, China, [2]Department of Criminal Science and Technology, Nanjing Police University, Nanjing, China

N^6-methyladenosine (m^{6}A), a well-characterized RNA modification, is involved in regulating multiple biological processes; however, genome-wide identification and functional characterization of the m^{6}A modification in legume plants, including soybean (*Glycine max* (L.) Merr.), remains lacking. In this study, we utilized bioinformatics tools to perform comprehensive analyses of molecular writer candidates associated with the RNA m^{6}A modification in soybean, characterizing their conserved domains, motifs, gene structures, promoters, and spatial expression patterns. Thirteen m^{6}A writer complex genes in soybean were identified, which were assigned to four families: MT-A70, WTAP, VIR, and HAKAI. It also can be identified that multiple cis elements in the promoters of these genes, which were classified into five distinct groups, including elements responsive to light, phytohormone regulation, environmental stress, development, and others, suggesting that these genes may modulate various cellular and physiological processes in plants. Importantly, the enzymatic activities of two identified m^{6}A writers, GmMTA1 and GmMTA2, were confirmed *in vitro*. Furthermore, we analyzed the expression patterns of the *GmMTA*s and *GmMTB*s under different abiotic stresses, revealing their potential involvement in stress tolerance, especially in the response to alkalinity or darkness. Overexpressing *GmMTA2* and *GmMTB1* in soybean altered the tolerance of the plants to alkalinity and long-term darkness, further confirming their effect on the stress response. Collectively, our findings identified the RNA m^{6}A writer candidates in leguminous plants and highlighted the potential roles of *GmMTA*s and *GmMTB*s in the response to abiotic stress in soybean.

KEYWORDS

m^{6}A, soybean, RNA methylation, abiotic stress, MTA, MTB

Introduction

Over 150 distinct chemical modifications of eukaryotic RNA molecules have been identified, including methylation, acetylation, and glycosylation (Delaunay et al., 2023). Many exist in noncoding RNAs, particularly transfer RNAs and ribosomal RNAs. Massage RNA (mRNA) can also carry several base modifications, such as N^6-methyladenosine (m^6A), N^1-methyladenosine, 5-methylcytidine, N^4-acetylcytidine, N^7-methylguanosine, and pseudouridine (Frye et al., 2016, 2018). These chemical modifications influence gene expression by regulating the structure, splicing, transport, stability, and translation efficiency of the target transcripts (Roundtree et al., 2017). Of all known modifications, m^6A is the most abundant in mRNA (Jia et al., 2013). It is installed by a "writer" protein complex and dynamically removed by "eraser" proteins in the nucleus (Shi et al., 2019). Often, "reader" proteins are responsible for decoding the m^6A signature (Han et al., 2021).

In plants, m^6A modifications in mRNA are decorated by a methyltransferase complex comprising MRNA ADENOSINE METHYLASE A (MTA; the ortholog of human METHYLTRANSFERASE-LIKE 3 (METTL3), MRNA ADENOSINE METHYLASE B (MTB; the ortholog of human METTL14), FKBP12 INTERACTING PROTEIN 37 kDa (FIP37), VIRILIZER (VIR), and the E3 ubiquitin ligase HAKAI (Zhong et al., 2008; Shen et al., 2016). Mutations of *MTA*, *MTB*, *FIP37*, or *VIR* in Arabidopsis (*Arabidopsis thaliana* (L.) Heynh.) are embryo-lethal, indicating that m^6A RNA modifications are essential for plant development (Vespa et al., 2004; Zhong et al., 2008; Shen et al., 2016; Růžička et al., 2017). The absence of HAKAI decreases the abundance of m^6A modifications but does not result in obvious growth defects (Růžička et al., 2017). In cotton (*Gossypium hirsutum* (L.)), m^6A modifications enhance the stability of *GhMYB44* mRNA, contributing to fiber elongation and secondary cell wall thickening (Xing et al., 2023). m^6A modifications in plants are also involved in the stress response. In Arabidopsis, transcripts involved in the salt and osmotic stress responses showed an increased abundance of m^6A modifications when the plants were grown under high salinity (Anderson et al., 2018). In apple (*Malus domestica* (Suckow) Borkh.), MdMTA-mediated m^6A modifications improved drought tolerance by promoting the mRNA stability and translation efficiency of genes associated with lignin deposition and oxidative stress (Hou et al., 2022). The overexpression of the m^6A reader *Malus hupehensis* YTH-domain family protein 2 (MhYTP2) can elevate the mRNA stability of its target *M. domestica* allantoinase-like gene (*MdALN*) in apple (Guo et al., 2023). Together, this evidence highlights the significant roles of m^6A modifications in developmental regulation and stress tolerance in plants.

Soybean (*Glycine max* (L.) Merr.) is an economically important leguminous crop, as it contributes a huge proportion of the total global oilseed and biodiesel and provides vital protein and oil sources for human food and animal fodder (Qi and Lee, 2014; Kim et al., 2017). Soybean cultivation faces challenges from abiotic stresses, such as high or low temperatures and soil salinity or alkalinity. Over the course of its long evolutionary history, soybean has developed intricate strategies to withstand abiotic stresses, which involve modifications across multiple dimensions, including its metabolism, physiology, and transcriptome (Feng et al., 2021; Sheikh et al., 2024). One such abiotic stress, soil salinity, causes ion toxicity and osmotic stress, and thus induces the generation of reactive oxygen species (ROS) (van Zelm et al., 2020). Enzymatic antioxidants, such as superoxide dismutase (SOD), peroxidase (POD), and catalase (CAT), can remove ROS to reduce oxidative stress and protect plants from damage (Wang et al., 2016a; Fu et al., 2017). Additionally, ion channel proteins play a crucial role in the response to salt stress; for instance, Na^+/H^+ ANTIPORTER 1 (GmNHX1) sequesters Na^+ into the vacuole, thereby reducing its level in the cytoplasm (Yang et al., 2017). Similarly, CATION/PROTON EXCHANGER 1 (GmCHX1) facilitates the exclusion of Na^+ from leaf tissues, mitigating the toxic effects associated with excessive salt accumulation (Qu et al., 2022). Moreover, stress-inducible transcription factors (TFs), including NACs, bZIPs, MYBs, and WRKYs (Katiyar et al., 2012; Ji et al., 2013; Zhu et al., 2014; Phukan et al., 2016; Lim et al., 2022), enhance plant stress tolerance by regulating the expression of target genes during the stress response. Gene expression is also modulated by epigenetic modifications, including m^6A, which play important roles in plant responses to abiotic stress; however, studies on the involvement of m^6A in the soybean response to abiotic stress are limited, and the key components involved have yet to be identified.

In this study, we employed genome-wide analyses to identify the m^6A writer proteins in soybean and analyzed the gene structure and evolutionary aspects of each family member. We particularly focused on elucidating the expression patterns and subcellular localization of the core members of the m^6A writer complex, the GmMTAs and GmMTBs, which were determined to be involved in abiotic stress responses. Our study provides a foundation for further exploring the function of the m^6A modification in soybean.

Materials and methods

Plant materials and stress treatments

Soybean (*Glycine max*, Williams 82) seeds were sterilized using chlorine and then planted on moist vermiculite. The plants were cultivated in a growth chamber under long-day conditions (16 h light of 100 μmol m^{-2} s^{-1} intensity provided by white LED lamps at 25°C and 8 h dark at 23°C, 70% relative humidity). 15-day-old seedlings were used for the different abiotic stress treatment. For the cold or heat treatments, seedlings were grown at 8°C or 42°C for 24 h, respectively. For the drought stress treatment, soybean seedlings were transferred to Hoagland liquid culture containing 20% polyethylene glycol (PEG) and incubated for 1 day. For the salinity or alkalinity stress treatments, soybean plants were grown in Hoagland liquid culture containing 150 mM NaCl or 100 mM $NaHCO_3$, respectively, for 1 day. For the darkness treatment, soybean seedlings were grown in Hoagland liquid culture under the dark condition for 3 days. Following these abiotic stress treatments, the leave and roots samples were collected separately and frozen at -80°C for subsequent experiments.

Soybean leaf transient transformation

To generate transgenic soybean lines overexpressing *GmMTA2* or *GmMTB1*, a modified soybean transient transformation was utilized based on a previously study (Wang et al., 2023a). In brief, the transformed *Agrobacterium* cells were incubated with infiltration buffer (OD_{600} = 1) containing 10 mM MES (pH 5.6) and 200 μM acetosyringone (Sigma) and infiltrated into the lower epidermis of the leaves from 7-day-old soybean seedlings by employing a vacuum pump until the leaves were thoroughly wetted. Following a recovery period of one day under the continuous darkness, the transformed soybean seedlings were transferred to the normal growth condition for another five days before being subjected to the stress treatment. For alkalinity treatment, the seedlings were grown in Hoagland liquid culture containing 100 mM $NaHCO_3$ for 36 hours. Histochemical detection of hydrogen peroxide (H_2O_2) and superoxide anion ($O^{2\cdot-}$) and the enzymatic activity measurement of CAT, POD, and SOD from the infiltrated leaves were performed according to previous studies (Cheng et al., 2019; Pi et al., 2023). For darkness treatment, the seedlings were grown under the continuous darkness for 10 days. Chlorophyll content of soybean leaves from each sample was measured according to a previous description (Li et al., 2024).

RNA extraction, cloning, and quantitative RT-PCR analyses

Total RNA was isolated from 15-day-old soybean leaves and cDNA was prepared according to previous studies (Anderson et al., 2018; Gao et al., 2023; Wang et al., 2023b). The following primers were used for cloning: for *GmMTA1* (Glyma.07G067100), Cp994 and Cp995; for *GmMTA2* (Glyma.16G033100), Cp1293 and Cp1294, for *GmMTB1* (Glyma.10G232300), Cp988 and Cp989; for *GmMTB2* (Glyma.20G161800), Cp1291 and Cp1292 (Supplementary Table S3). *GmMTA1*, *GmMTA2*, *GmMTB1*, and *GmMTB2* were cloned into pBA002-flag-HA-StrepII (VN21) and pXCS-YFP (V36) (Chen and Witte, 2020), respectively, for the protein purification and subcellular localization analysis. The VN21 was generated by introducing a 135-bp fragment encoding the flag-HA-StrepII tag using the primer pair Cp795/Cp796 into pBA002 vector (Kost et al., 2008).

To analyze mRNA abundance, reverse transcription quantitative PCR (RT-qPCR) was performed with QuantStudio 1 (Thermo Fisher Scientific) using Hieff qPCR SYBR Green Master Mix (Yeasen Biotechnology) according to the previous description (Wang et al., 2023b). Total RNA from each sample was isolated using the RNA isolater Total RNA Extraction Reagent (Vazyme Biotech Co., Ltd). The quantity and concentration of the resulted RNA were evaluated using a Thermo Scientific NanoDrop™ spectrophotometer. A total of 800 ng of RNA was used to synthesize the first-strand cDNA using an Oligo (dt) primer. Transcript abundance of *GmMTA1*, *GmMTA2*, *GmMTB1*, and *GmMTB2* was analyzed by employing the primer pairs Cp633/Cp634 or Cp631/Cp632, Cp637/Cp638, and Cp635/Cp636, respectively. *GmF-BOX* (Glyma.12G051100) was amplified with the primer pairs Cp363/Cp364 as an internal reference gene in mRNA. The calculation was based on the $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen, 2001). All primer sequences are detailed in Supplementary Table S3.

Phylogenetic and characteristics analyses of the components m^{6}A writer proteins in soybean

The protein sequences of Arabidopsis m^{6}A writers were obtained from the TAIR database (https://www.arabidopsis.org) and used as reference sequences. Subsequently, the m^{6}A writer protein sequences from *Glycine max*, *Glycine soja, Phaseolus vulgaris, Medicago truncatula*, and *Lotus japonicus* were identified by using the BlastP method from Phytozome database (https://phytozome-next.jgi.doe.gov). Multiple sequence alignment was performed with the ClustalW method (Larkin et al., 2007). To compare evolutionary relationships, the m^{6}A writer protein sequences from *A. thaliana*, *G. max*, *G. soja*, *P. vulgaris*, *M. truncatula*, and *L. japonicus* were used to construct a phylogenetic tree using MEGA11 (Tamura et al., 2021) with the Neighbor-Joining (NJ) method and 1000 bootstrap replications. Thereafter, the phylogenetic tree was visualized using ChiPlot (https://www.chiplot.online/) (Xie et al., 2023).

Analyses of gene structure, conserved motifs, collinearity relationship, and cis-elements analyses of the m^{6}A writer genes

Gene structures were visualized using the TBtools software with GFF files provided as input. Conserved motifs within the m^{6}A writer protein sequences were analyzed using the MEME online tool (https://meme-suite.org/meme/tools/meme). The parameters were set as following: the site distribution was designated as 'any number of repetitions' (anr), the number of motifs was specified as 10, and all other optional parameters were kept at their default settings. Results of the conserved domains were visualized by using TBtools. To investigate the collinearity relationships among m^{6}A writer genes in *G. max*, the One Step MCScanX-Super Fast program integrated into TBtools was employed. For cis-elements analysis, 2000-bp region upstream of the start codons of m^{6}A writer genes were obtained from the Phytozome database. Subsequently, the promoter sequences were submitted to the PlantCARE database (https://bioinformatics.psb.ugent.be/webtools/plantcare/html/). Prediction from the PlantCARE database were visualized using TBtools (Chen et al., 2020).

Tissue-specific expression of the m^{6}A writer proteins in soybean

To analyze the tissue-specific expression of m^{6}A writer proteins, transcriptome sequencing data were obtained from the SoyOmics

database (Liu et al., 2023) for various soybean tissues, including cotyledon, stem, leaf bud, leaf, flower, seed, shoot, and root. These data were graphically represented and visualized using the ChiPlot.

Subcellular localization

GmMTAs-eYFP or GmMTBs-eYFP was transiently co-expressed with the nucleus marker protein RFP-H2B (RFP fused to histone 2B) (Martin et al., 2009) in *Nicotiana benthamiana* leaves for 5 days. The samples were analyzed using a ZEISS LSM 980 with Airyscan2 microscope equipped with an HC PLAPO CS2 40 × 1.0 water immersion objective (ZEISS Microsystems) according to the previous description (Wang et al., 2023b).

Protein purification and enzymatic activity measurement

Recombinant soybean GmMTA1 and GmMTA2 were affinity purified after transient expression in *N. benthamiana* as described before (Chen et al., 2016; Baccolini and Witte, 2019). Purified protein was quantified by employing the Bradford reagent from Tiangen with bovine serum albumin (BSA) as the standard.

To assess the enzymatic activity, 0.15 nmol of recombinant GmMTA1 or GmMTA2 was incubated with a 50 µl substrate solution, containing 0.8 mM S-adenosylmethionine (SAM; Sigma), 0.15 nmol RNA probe (UACACUCGAUCUGGACUAAAGC UGCUC, synthesized by Genscript), 80 mM KCl, 1.5 mM $MgCl_2$, 0.2 U μl^{-1} RNasin, 10 mM dithiothreitol (DTT), 4% glycerol and 15 mM HEPES (pH 7.9), at 28°C for 1 h. Then the reaction was terminated by incubating at 95°C for 15min. The resulting RNA product (800 ng) were fully digested into single nucleosides in a 50µl reaction buffer containing 10 mM Tris–HCl, pH 7.9, 1 mM $MgCl_2$, 0.1 mg mL^{-1} BSA, 0.4 units benzonase (Sigma-Aldrich), 0.004 units phosphodiesterase I (Sigma-Aldrich) and 0.04 units shrimp alkaline phosphatase (NEB) according to previous description with minor revision (Chen et al., 2018; Chen and Witte, 2020; Gao et al., 2023; Wang et al., 2023b). After incubation at 37°C for 10 h, enzyme reaction was terminated and the sample was filtered by an ultrafiltration tube (3 kDa cutoff; Pall). 2µl aliquots were analyzed by an Agilent 1290 HPLC system coupled with a Sciex 6500 Qtrap mass spectrometer. The following mass transitions were monitored: m/z 268.1 to 136 (A, adenosine); m/z 282.12 to 150 (m^6A, N6-methyladenosine). Standard solutions of A: 1, 5, 25 50, 100, 200, 400, 2000 and 10 000 ng/ml; m^6A: 0.1, 0.5, 2.5, 5, 10, 20, 40, 200, and 1000 ng/ml were used for quantification. The ratio of m^6A to A were calculated based on the calibrated concentrations.

Statistics

Statistical analysis was performed by GraphPad Prism 9.5.1 software. Statistical methods and sample sizes are shown in the figure legends. All replicates are biological replicates or experimental replicates.

Results

Genome-wide identification and evolutionary analysis of mRNA m^6A writer genes in legume plants

To identify the m^6A writer candidate genes in legume plants, we used the sequences of the known writer proteins MTA, MTB, MTC, FIP37, VIR, and HAKAI from the model plant Arabidopsis as queries in BLASTp searches against the genomes of soybean (*G. max*), wild soybean (*G. soja* Siebold & Zucc.), common bean (*Phaseolus vulgaris* L.), *Medicago truncatula* Gaertn., and *Lotus japonicus* (Regel) K. Larsen in the Phytozome V13 database (https://phytozome-next.jgi.doe.gov/blast-search). We identified thirteen candidate m^6A writer homologs in *G. max*, thirteen in *G. soja*, six in *P. vulgaris*, eight in *M. truncatula*, and six in *L. japonicus* (Figure 1; Supplementary Table S1). All identified proteins were divided into four families (MT-A70, WTAP, VIR, and HAKAI) according to their topological structure. At least one homolog for each family was identified in each of the plant species, with the MT-A70 family representing the most identified (22 candidates in total) and the least identified for the VIR family (eight in total). Among the five analyzed legumes, the *G. max* and *G. soja* genomes contained the highest number of m^6A writer candidate genes (Supplementary Table S1). In *G. max*, five candidate genes were identified in the MT-A70 family, including *Glyma.14G077000*, *Glyma.10G232300*, *Glyma.20G161800*, *Glyma.07G067100*, and *Glyma.16G033100*. Four candidate genes were identified for the WTAP family, and two each were revealed for the VIR or HAKAI families (Figure 1; Table 1). The members of each writer protein family were named sequentially according to their order on the chromosomes (Table 1).

We identified similar characteristics in the candidate proteins from each family in *G. max*. The lengths of the MT-A70 family candidate proteins ranged from 428 to 1102 amino acids. The predicted molecular weights were 48.88 to 121.73 kDa, with theoretical isoelectric points (pIs) in the range of 5.95 to 7.14. The two VIR family candidates had the longest sequences, which were 2230 and 2174 aa with molecular weights of 246.01 and 239.33 kDa and pIs of 5.36 and 5.35. The HAKAI and WTAP candidates were shorter, ranging from 439 to 440 aa and 343 to 354 aa, respectively. Their molecular weights ranged from 47.89 to 47.94 kDa (HAKAI family) and 38.34 to 40.34 kDa (WTAP family), and their pIs were 6.15 to 6.2 and 5.04 to 5.75, respectively.

Conserved motifs and gene exon–intron structures of the mRNA m^6A writer genes

We constructed a phylogenetic tree using the Neighbor-Joining method to reconstruct the evolutionary relationships among all writer candidates from *G. max* (Figure 2). The result was consistent with that of the phylogenetic analysis constructed using the proteins from all five legume plants and Arabidopsis (Figure 1). We employed a motif analysis using the MEME program to identify the conserved

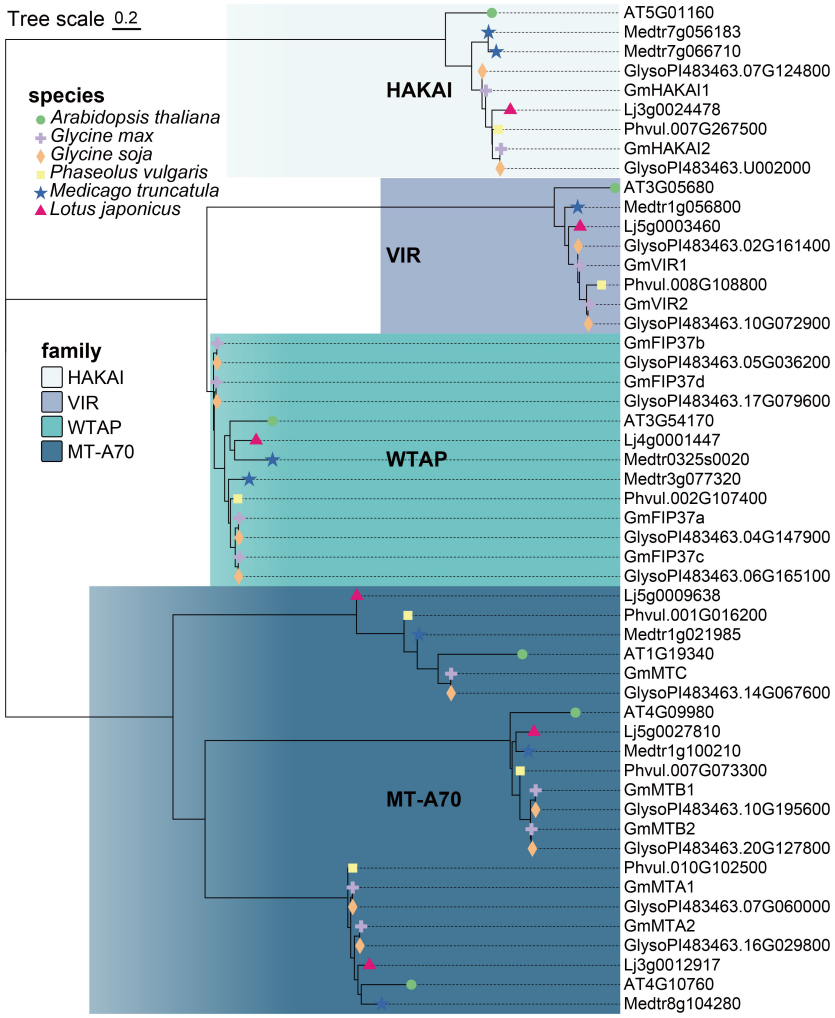


FIGURE 1
Phylogenetic analysis of m^{6}A writer proteins in *Arabidopsis thaliana*, *Glycine max*, *Glycine soja*, *Phaseolus vulgaris*, *Medicago truncatula*, and *Lotus japonicus*. The phylogenetic tree was constructed using MEGA11 software with the Neighbor-Joining algorithm and 1,000 bootstrap replicates.

motifs present within the m^{6}A writer candidates in *G. max*. In total, 10 distinct and highly conserved motifs were predicted (Figure 2A; Supplementary Table S2). Each candidate protein contained 2–6 motifs except GmMTC (Figure 2A), hinting that most are likely m^{6}A writer proteins. Motif 4, motif 7, and motif 10 were common to many members, suggesting that they may be important for methyltransferase activity. Additionally, motif 1, motif 2, and motif 3 were unique to the WTAP/FIP37 subfamily, while motif 5 and motif 8 exclusively existed in the MT-A70 subfamily. The homologous proteins of each family share identical conserved motifs, suggesting they may be functionally redundant.

We analyzed the exon–intron distribution to investigate the genetic structural diversity. Notably, m^{6}A writer candidates from each subfamily had a similar exon–intron pattern, although across all writer candidates the exon number varied substantially, from 3 to 28. The two *GmVIR*s had the most exons (27 and 28 exons), while two *GmHAKAI*s had three exons each. Seven and six exons, respectively, were identified in the core methyltransferase genes, *GmMTA*s and *GmMTB*s. Generally, closely related candidate writers tended to have similar conserved motifs (Figure 2A) and exon–intron structure patterns (Figure 2B), suggesting their relative conservation during the evolutionary process and thereby substantiating the accuracy of the clustering analysis presented in Figure 1.

Chromosomal distribution of mRNA m^{6}A writer genes

We investigated the chromosomal distribution of the m^{6}A writer genes in soybean and identified associated gene duplication events. In total, 13 genes were randomly distributed on 11 of the 20 chromosomes of *G. max* (Figure 3). Chromosomes 7 and 10 each possess two m^{6}A writer genes, whereas chromosomes 2, 4, 5, 6, 14, 16, 17, 18, and 20 each contain one such gene. Different gene replication events, such as tandem duplication and fragment duplication, occur in plant genomes, resulting in the expansion of gene families (Cannon et al., 2004). The collinearity analysis

10.3389/fpls.2024.1446591

TABLE 1 Characteristics of predicted m^{6}A writer candidate genes in *Glycine max*.

Family	Gene name	Gene ID (Phytozome)	Amino acid length	Isoelectric point	Molecular weight (kDa)	Subcellular localization prediction	Orthologous gene ID in *A. thaliana*
MT-A70	*GmMTA1*	Glyma.07G067100	762	6.22	84.61	nucleus	AT4G10760
	GmMTA2	Glyma.16G033100	761	5.95	84.26	nucleus	
	GmMTB1	Glyma.10G232300	1102	6.65	121.73	nucleus	AT4G09980
	GmMTB2	Glyma.20G161800	1098	6.76	121.47	nucleus	
	GmMTC	Glyma.14G077000	428	7.14	48.88	nucleus	AT1G19340
WTAP	*GmFIP37a*	Glyma.04G186400	354	5.75	40.34	nucleus	AT3G54170
	GmFIP37b	Glyma.05G040200	339	5.12	38.34	nucleus	
	GmFIP37c	Glyma.06G179400	343	5.59	39.17	nucleus	
	GmFIP37d	Glyma.17G086600	343	5.04	38.75	nucleus	
VIR	*GmVIR1*	Glyma.02G195600	2230	5.36	246.01	nucleus	AT3G05680
	GmVIR2	Glyma.10G082100	2174	5.35	239.33	nucleus	
HAKAI	*GmHAKAI1*	Glyma.07G144300	439	6.2	47.89	nucleus	AT5G01160
	GmHAKAI2	Glyma.18G195500	440	6.15	47.94	nucleus	

revealed no tandem duplication between the m^{6}A writer genes; however, nine pairs arising from fragment duplications (*GmMTA1*/*GmMTA2*, *GmMTB1*/*GmMTB2*, *GmMTC*/*Glyma.17G248801*, *GmVIR1*/*GmVIR2*, *GmHAKAI1*/*GmHAKAI2*, *GmFIP37a*/*GmFIP37b*, *GmFIP37b*/*GmFIP37c*, *GmFIP37c*/*GmFIP37d*, and *GmFIP37a*/*GmFIP37d*) were observed, showing that members from each m^{6}A writer family originated from gene duplication in soybean. A duplication partner from one such pair, *Glyma.17G248801*, is an uncharacterized gene without an MT-A70 domain, suggesting that it possesses an unknown function other than m^{6}A methyltransferase activity.

Spatial expression patterns and subcellular localization of the mRNA m^{6}A writer genes and proteins

Employing the publicly available data from the SoyOmics website (https://ngdc.cncb.ac.cn/soyomics/index), we examined the expression profiles of all m^{6}A writer genes across various tissues from *G. max*, including the cotyledon, stem, leaf bud, leaf, flower, pod_seed (pod plus seed), pod, seed, shoot meristem, and root. Our analyses revealed distinct expression patterns among these genes across different tissue types (Supplementary Figure S1). Notably, we observed a pronounced upregulation of the expression of m^{6}A writer genes in the leaf bud, flower, and shoot, indicating their potential involvement in orchestrating floral development. By contrast, we were surprised to discover relatively low expression levels of these genes in the leaves.

Furthermore, we used WoLF PSORT (https://wolfpsort.hgc.jp/) to predict the subcellular localization of all members, which were all predicted to be localized in the nucleus (Table 1). To test this prediction, we fused the full-length amino acid sequences of the four core members, GmMTA1, GmMTA2, GmMTB1, and GmMTB2, with a C-terminal yellow fluorescent protein (YFP) tag each and then independently transiently co-expressed them with the nucleus marker (H2B) fused with mCherry (RFP) in *N. benthamiana* leaves. The results showed that all four proteins were indeed localized to the nucleus (Figure 4), which is consistent with the prediction.

Methyltransferase activity analysis of the GmMTAs

In mammals, an MTase complex comprising METTL3 and METTL14 efficiently catalyzes the transfer of a methyl group to m^{6}A on RNA. METTL3 primarily functions as the catalytic core, while METTL14 acts as an RNA-binding platform (Wang et al., 2016b). To investigate the enzymatic activities of GmMTAs, which are the homologs of mammalian METTL3 in soybean, we independently expressed *GmMTA1* and *GmMTA2* recombinantly in *N. benthamiana* and performed methylation assays with the purified proteins using an unmodified RNA probe and SAM as the methyl donor (Figure 5A). Subsequently, the RNA product was digested into nucleosides and the abundance of m^{6}A was analyzed using liquid chromatography–tandem mass spectrometry (LC-MS/MS). Remarkably, our findings revealed a significant increase in m^{6}A levels upon adding GmMTA1 or GmMTA2 proteins compared with the control. These experimental results strongly support our hypothesis that GmMTA1 and GmMTA2 possess similar methyltransferase activities, indicating their ability to catalyze mRNA m^{6}A modifications (Figure 5).

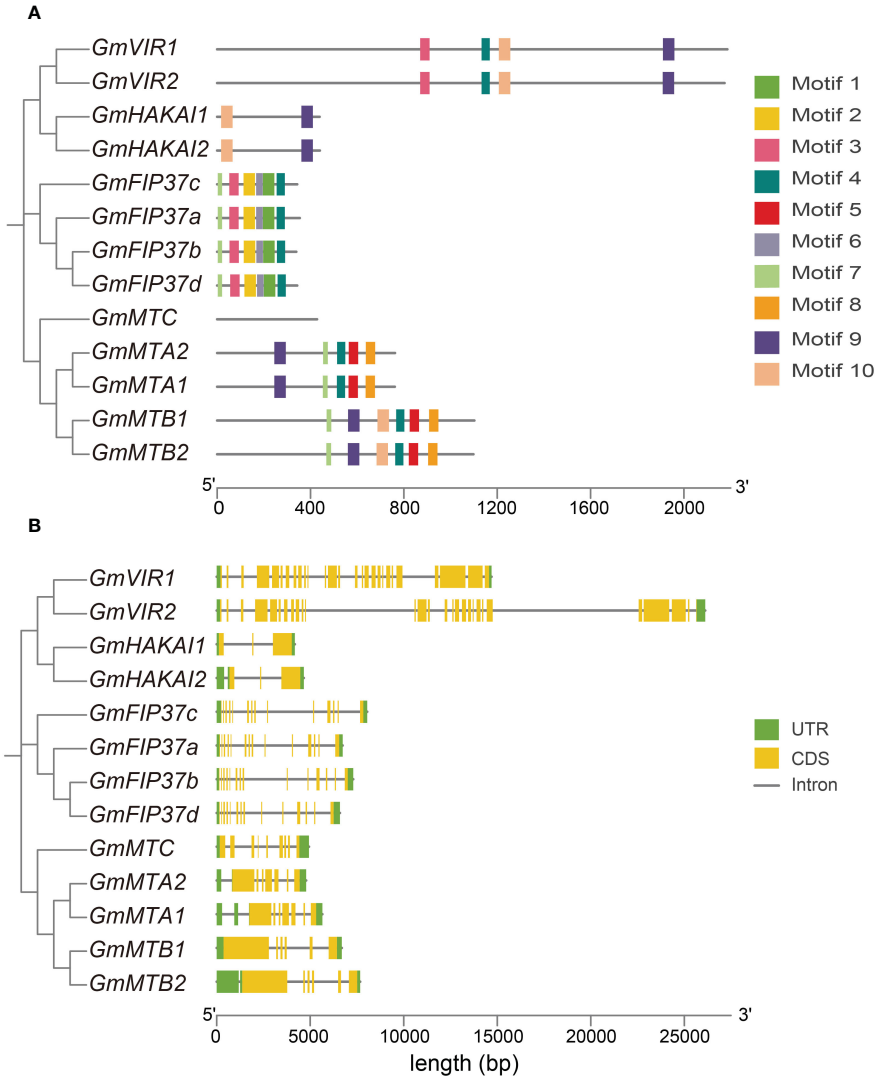


FIGURE 2
The conserved motifs and gene structure of m^{6}A writer genes in *G max*. **(A)** Phylogenetic analysis of m^{6}A writer candidates from *G max* and the organization and distribution of the conserved motifs in the m^{6}A writer genes. **(B)** Phylogenetic analysis of m^{6}A writer candidates from *G max* and the exon–intron structures of the m^{6}A writer genes. Untranslated regions (UTRs) are represented by green boxes, coding sequences (CDSs) are represented by yellow boxes, and introns are represented by gray lines.

Cis-element analyses of mRNA m^{6}A writer genes

To gain deeper insights into the transcriptional regulatory activity of RNA m^{6}A writer genes, we predicted the cis elements within the 2000-bp promoter regions of all candidates using the PlantCARE web server (http://bioinformatics.psb.ugent.be/webtools/). The identified cis-acting elements were involved in 25 functional categories, which could be classified into five groups: light-responsive elements, phytohormone-responsive elements, environmental stress-related elements, development-responsive elements, and other elements (Figures 6A, B). Remarkably, light-responsive elements could be identified in all promoters. Among the phytohormone-responsive elements, those involved in the abscisic acid (ABA) response and methyl jasmonate (MeJA) response were the most abundant. The most frequent environmental stress-related elements were anaerobic induction, followed by the low temperature-responsive ones. Furthermore, cis elements related to the developmental response were also identified (Figure 6B). The detection of light-responsive elements in all promoters (Figure 6C) suggests that the expression levels of m^{6}A writer genes are probably influenced by light, which may affect the development and environmental stimuli responses of soybean.

Expression patterns of *GmMTA*s and *GmMTB*s in response to abiotic stress

The results of the cis-element analysis suggest that m^{6}A RNA modifications in soybean play crucial roles in the response to different abiotic stresses. We therefore analyzed the expression pattern of the m^{6}A writer components *GmMTA1, GmMTA2,*

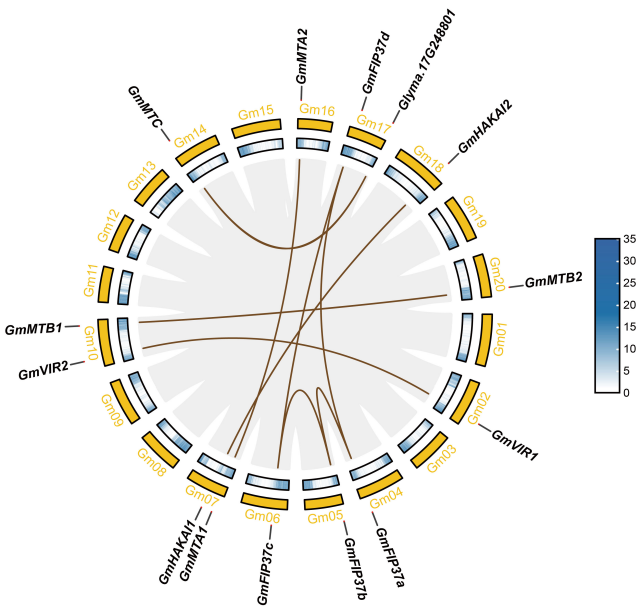


FIGURE 3
Collinearity analysis of the m^{6}A writer genes in soybean. The tan lines represent collinear pairs of m^{6}A writer genes, and the gray lines represent the collinearity results of the soybean genome.

GmMTB1, and *GmMTB2* in the leaf and root of 15-day-old soybean plants grown under normal conditions or different stresses, including cold, heat, drought, salinity, alkalinity, or darkness, over 24 h. Generally, the abiotic stress treatments significantly altered the expression patterns of the *GmMTA*s and *GmMTB*s compared with the control (Figure 7; Supplementary Figure S2). Notably, each gene exhibited distinct responses to different stress conditions in the leaf; for instance, *GmMTA1* was initially suppressed after 2 h of cold treatment but induced after 6 h, whereas the expression level of *GmMTA2* remained unchanged until 12 h. By contrast, the

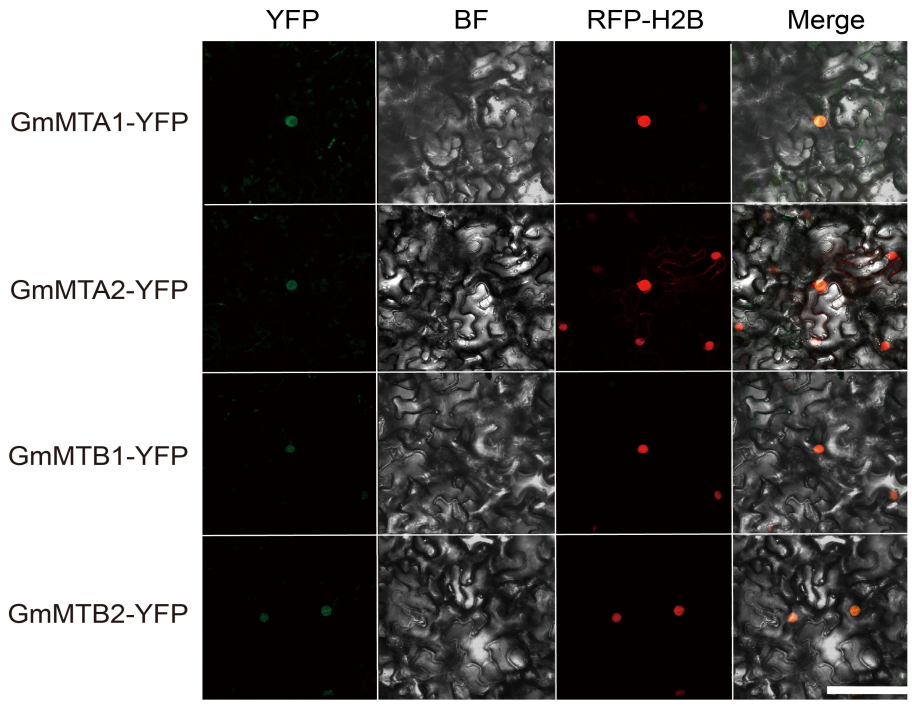


FIGURE 4
Subcellular localization analysis of GmMTAs and GmMTBs in *N. benthamiana* leaves. RFP-H2B was used as a nuclear marker. Bar, 100 μM. BF, brightfield.

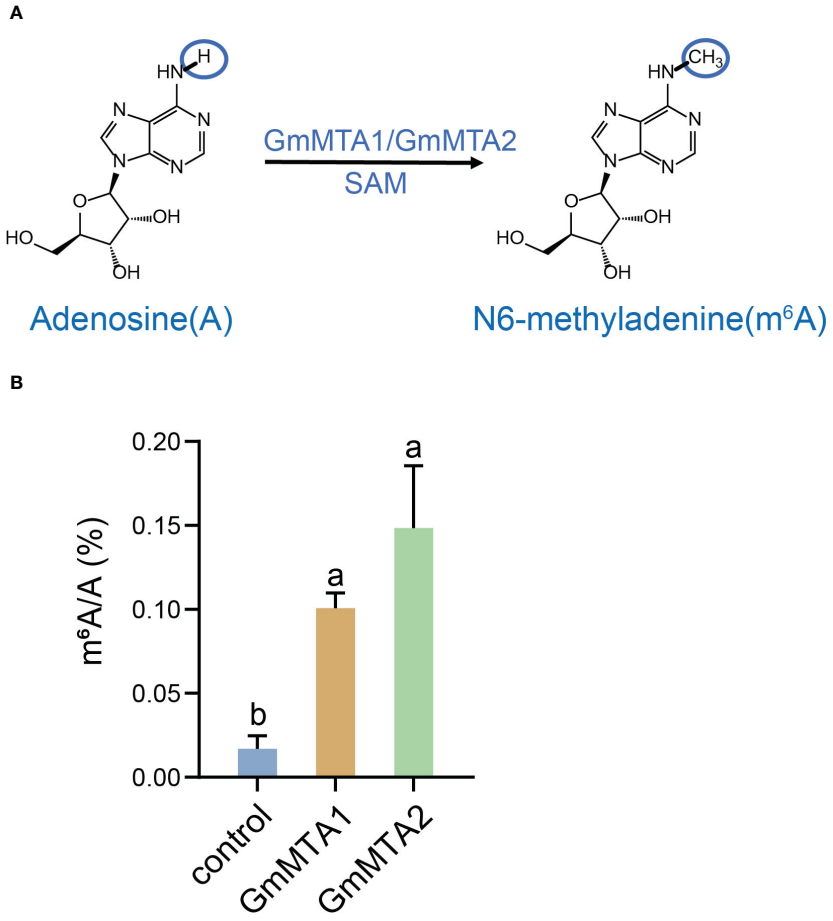


FIGURE 5
Biochemical analyses of GmMTA1 and GmMTA2. **(A)** A proposed reaction mechanism of adenosine methylation producing the N^6-methyladenosine (m^6A) RNA modification performed by GmMTAs in soybean. **(B)** *In vitro* methylation assay of GmMTA1 and GmMTA2. Error bars represent SD (n = 3). Different letters indicate significant differences at $P < 0.05$.

expression levels of *GmMTB1* and *GmMTB2* were both enhanced during the cold treatment (Figure 7B).

In addition, we also observed that *GmMTA*s and *GmMTB*s exhibit different response times to various stresses in the leaf. The expression of all four genes was increased after 12 to 24 h of alkalinity stress ($NaHCO_3$) treatment (Figure 7E), while darkness induced these changes after 6 to 12 h (Figure 7F). These data provide evidence that GmMTA- and GmMTB-mediated m^6A modifications generally participate in the abiotic stress response in soybean. Moreover, our findings also suggest that *GmMTA2* and *GmMTB1* are the dominant genes involved in the cellular response to these environmental stressors.

Overexpression of *GmMTA2* and *GmMTB1* in soybean enhances plant tolerance to alkalinity and darkness

To further verify the roles of GmMTAs and GmMTBs in the response to abiotic stress, we overexpressed *GmMTA2* and *GmMTB1* in the leaves of 7-day-old wild-type soybean plants, respectively, which were then subjected to alkalinity or darkness treatment. Reverse transcription quantitative PCR (RT-qPCR) and immunoblot analyses revealed that *pBA002-GmMTA2-HA* (*GmMTA2-OE*) and *pBA002-GmMTB1-HA* (*GmMTB1-OE*) were all overexpressed in the transformed soybean (Figure 8). Compared with the control expressing the empty pBA002 vector (EV), the leaves overexpressing *GmMTB1* but not *GmMTA2* exhibited significantly increased tolerance to the alkalinity treatment (Figure 9A). In addition, we detected the H_2O_2 content by Diaminobenzidine staining (Figure 9B) and measured the catalase (CAT), peroxidase (POD), and superoxide dismutase (SOD) activities (Figures 9C-E) in the leaves from each genotype after the $NaHCO_3$ treatment. These results confirmed that the leaves harboring *GmMTB1* overexpression were indeed more resistant to the alkalinity. During the darkness treatment, the leaves overexpressing *GmMTA2* showed the highest tolerance, while leaves overexpressing *GmMTB1* also performed better than the control EV leaves (Figure 10A). We confirmed these findings by measuring the chlorophyll contents (Figure 10B). These results, together with RT-qPCR analyses (Figure 7), demonstrated that GmMTB1 played a role in the tolerance to alkalinity treatment while both GmMTA2 and GmMTB1 were involved in dark stress.

10.3389/fpls.2024.1446591

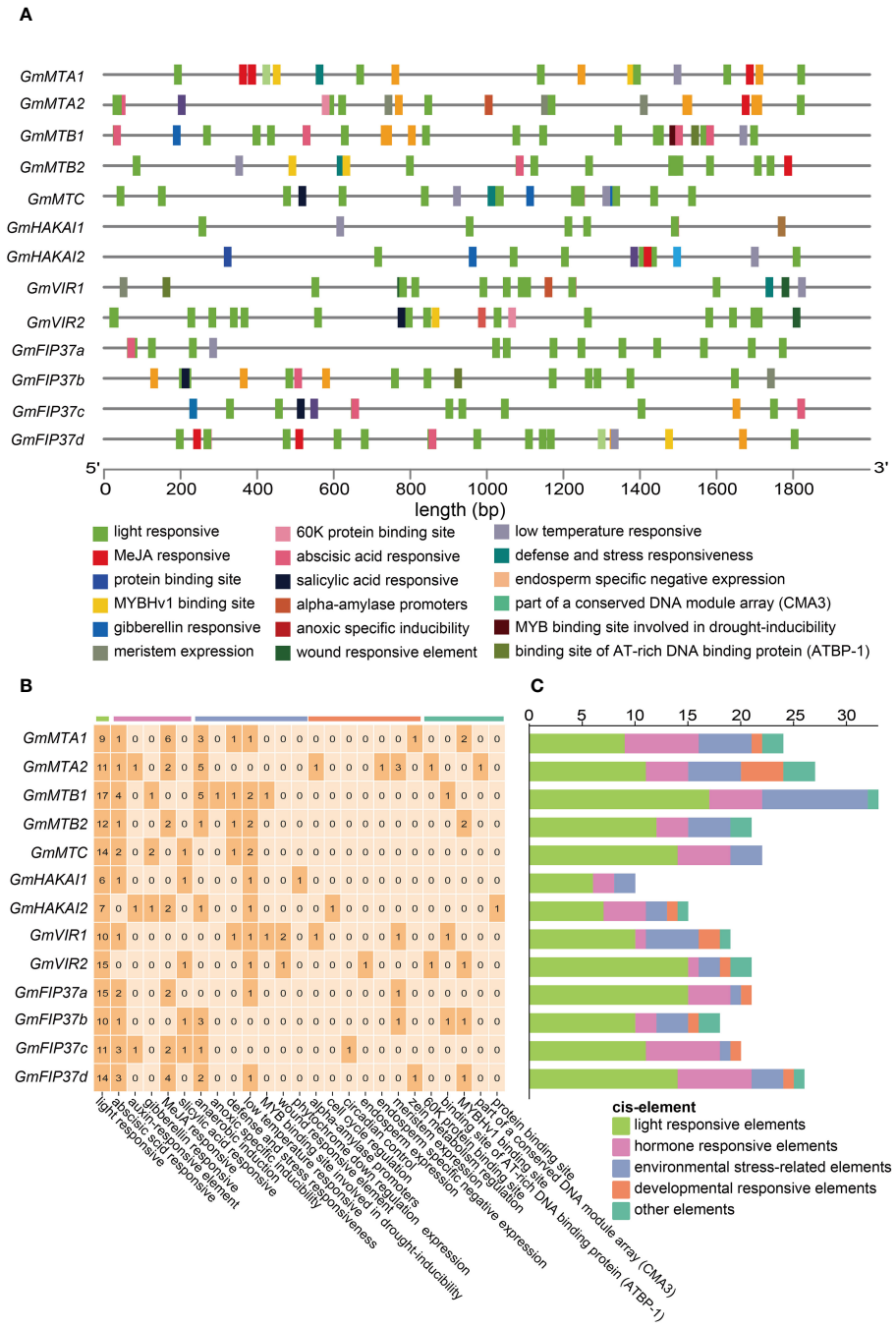


FIGURE 6
Cis elements in the promoters of m^{6}A writer genes. **(A)** The distribution of cis elements in the promoters of the m^{6}A writer genes. Different cis elements were depicted in different colored boxes. **(B)** The number of each category of cis element in m^{6}A writer genes. **(C)** Cis elements were classified into those responsive to light, phytohormones, development, environmental stress, and other regulated categories.

Discussion

Here, we identified and characterized the gene family members of the m^{6}A writer complex in legume plants. They were assigned to four families: MT-A70, WTAP, VIR, and HAKAI. In soybean, the MT-A70 family comprised five members, the WTAP subfamily had four members, and both the HAKAI and VIR families were represented by only two members each (Figure 1). Among these m^{6}A writer candidates, the enzymatic activities of GmMTA1 and GmMTA2 were confirmed in an *in vitro* assay (Figure 5). In addition, the gene structures of all candidates in *G. max* were analyzed, with similar structures and conserved motifs present among the members of each family. The gene structures varied between members of the different families; however, the VIR members exhibited more exons per gene (average of 27), while members of the HAKAI family had the fewest (only three exons per gene, on average; Figure 2). Upon analyzing the

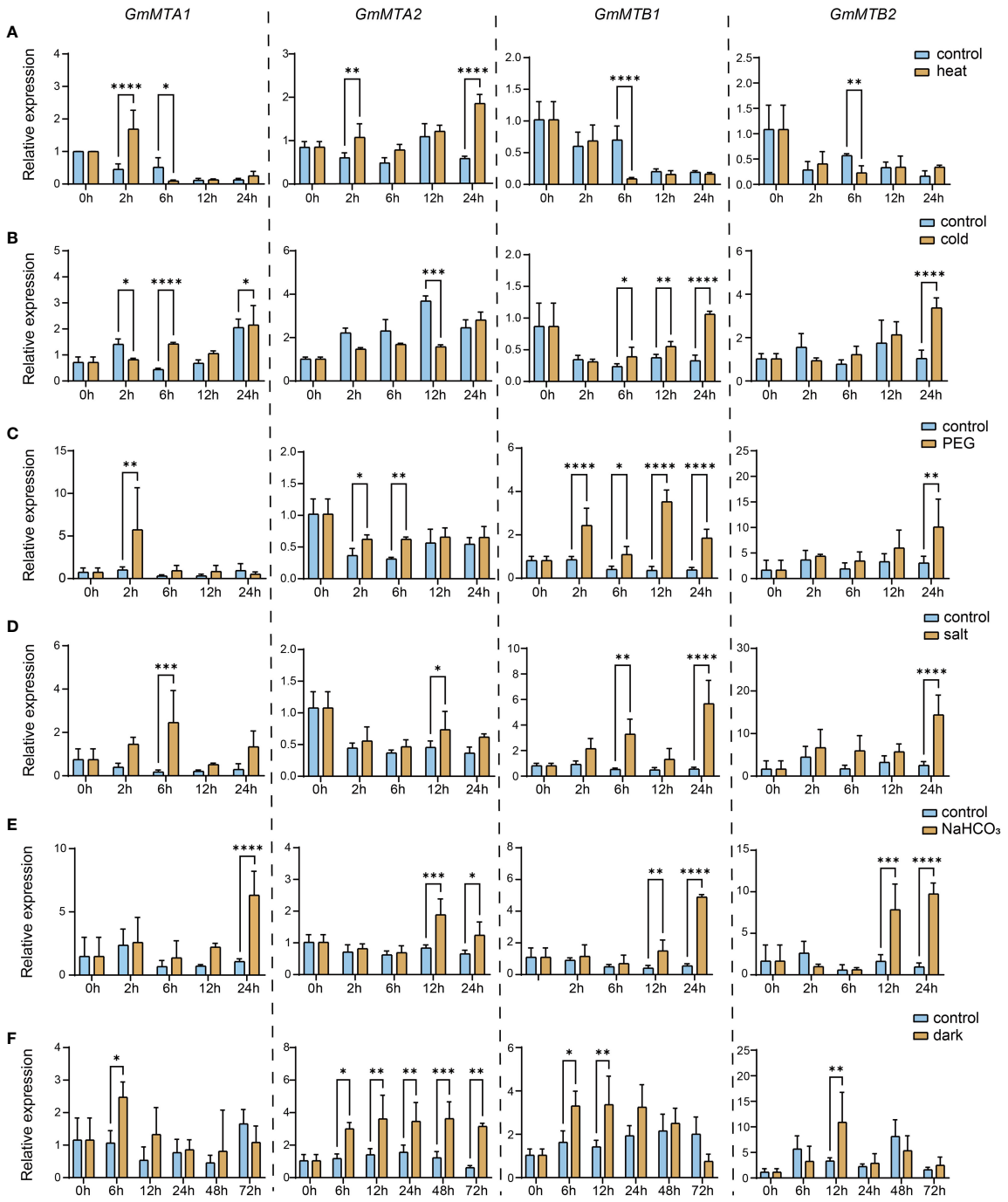


FIGURE 7
Relative expression levels of GmMTAs and GmMTBs in leaf under different abiotic stresses detected using reverse transcription quantitative PCR (RT-qPCR). 15-day-old soybean seedings were subjected to heat **(A)**, cold **(B)**, polyethylene glycol (PEG) **(C)**, salt **(D)**, alkalinity (NaHCO3) **(E)**, or darkness **(F)** treatments. GmF-BOX (Glyma.12G051100) was used as the internal control. Error bars represent SD (n = 3, *P ≤ 0.05; **P ≤ 0.01; ***P ≤ 0.001; ***P ≤ 0.0001).

promoter regions of the m^{6}A writer complex genes, several cis elements were identified. These cis elements were subsequently categorized into five functional clusters: responsive to light, phytohormone signaling (such as MeJA and ABA), environmental stress (such as low temperature or drought), development and other regulation (Figure 6). Simultaneously, we performed a subcellular localization analysis of all members of the MT-A70 family from *G. max*, revealing them all to be located within the nucleus, in line with our prediction (Table 1).

Abiotic stress reduces crop yields and can even kill the plant; therefore, it is critical to understand how soybean responds to environmental stresses, such as saline–alkali stress, cold stress, drought stress, and darkness (Hu et al., 2021; Zheng et al., 2021). There is mounting evidence that m^{6}A modifications play crucial roles

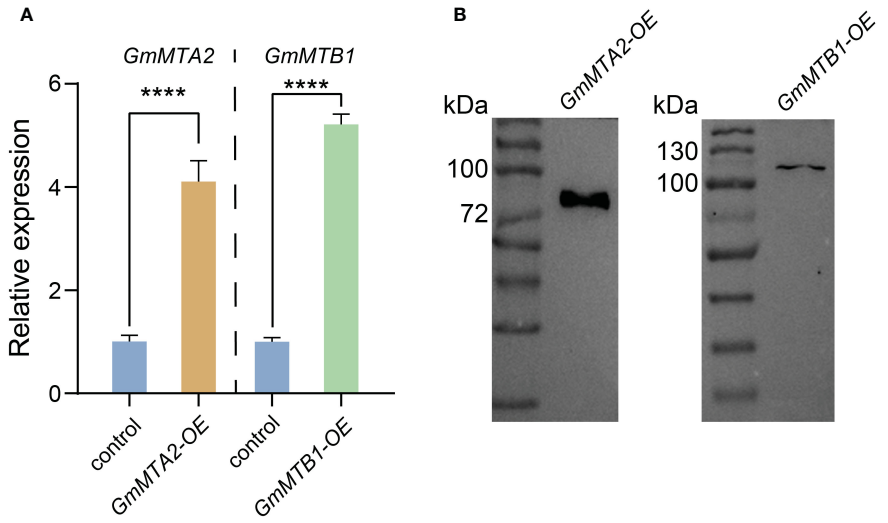


FIGURE 8
Confirmation of *GmMTA2* and *GmMTB1* overexpression in soybean. **(A)** Relative expression of *GmMTA2* and *GmMTB1* quantified using RT-qPCR in the wild type and the lines overexpressing *GmMTA2* (*GmMTA2-OE*) or *GmMTB1* (*GmMTB1-OE*). Error bars represent SD (n = 3, ****$P \leq 0.0001$). **(B)** Immunoblot analyses of leaf extracts from the soybean lines overexpressing *GmMTA2* or *GmMTB1*. An anti-HA antibody was used for detection. Control = plants expressing the empty vector.

in regulating plant responses to these stresses. In Arabidopsis, salt stress was reported to significantly affect the m^{6}A methylation levels on mRNA (Hu et al., 2021). Additionally, VIR-mediated m^{6}A methylation modulates ROS homeostasis by negatively regulating the mRNA stability of several negative regulators of the salt stress response, such as Arabidopsis NAC transcription factor (AtATAF1), GIGANTEA (AtGI), and glutathione S-transferaset U17 (AtGSTU17) (Zheng et al., 2021). In apple, MdMTA enhances lignin deposition and ROS scavenging under drought conditions. In poplar (*Populus trichocarpa*), plants overexpressing *PtrMTA* had a higher trichome density and a more developed root system (Lu et al., 2020). Moreover, they exhibited better drought tolerance (Lu et al., 2020). These findings show that the m^{6}A modification plays a pivotal role in the abiotic stress responses across diverse plant species, but our understanding of these mechanisms in soybean remains very limited. Here, we identified key m^{6}A component genes in soybean and analyzed their roles in different abiotic stress conditions, including cold, heat, drought, salinity, alkalinity, and darkness. Our data suggest that *GmMTA*s and *GmMTB*s, as core components of the m^{6}A writer complex, are altered in response to various stressors, especially induced upon drought, alkalinity, and darkness treatment; however, their expression patterns and response times differ in accordance with the specific stress conditions (Figure 7).

Leaf senescence has a crucial effect on crop quality and yield. It is an age-dependent process that can be regulated by several factors, including leaf age, phytohormones, temperature, and light. Upon entering the senescence stage, a leaf's cells undergo a sequential disorganization of cellular organelles, accompanied by systematic changes in metabolism and gene expression (Woo et al., 2019; Guo et al., 2021). Under darkness, the key transcription factors PHYTOCHROME-INTERACTING FACTOR 4/5 (PIF4/5) in the light signaling pathway are activated and regulate the expression of chlorophyll-catabolic genes through the ABA or ethylene pathways, accelerating chlorophyll degradation and promoting leaf yellowing and senescence (Sakuraba et al., 2014; Song et al., 2014). Long-term darkness has commonly been used as a tool to investigate the process of leaf senescence (Guo et al., 2021). In addition to transcriptional regulation, epigenetic modification also plays an important role in the regulation of leaf senescence. In Arabidopsis, the mutation of *MTA* resulted in a more pronounced aging phenotype than was observed in the wild type (Sheikh et al., 2024). The m^{6}A levels in the mRNA were shown to increase during a darkness treatment, preventing premature aging by destabilizing transcripts of age-related genes (Sheikh et al., 2024). We assessed the expression levels of the *GmMTA*s and *GmMTB*s under a darkness treatment and observed a significant upregulation after 6-12 h (Figure 7). In particular, *GmMTA2* showed continuous induction and maintained high expression levels after 72 h of treatment (Figure 7). Furthermore, we overexpressed *GmMTA2* and *GmMTB1* in soybean leaves, which resulted in a more prominent anti-aging phenotype and higher chlorophyll content (Figure 10). These findings suggest that soybean GmMTA2 and GmMTB1 play essential roles in dark-induced leaf senescence, providing additional evidence for m^{6}A involvement in crop or fruit ripening (Zhou et al., 2022). In animal studies related to m^{6}A, METTL3 (the human homolog of MTA) is also vital for slowing the aging of human mesenchymal stem cells (Wu et al., 2020); hence, further research into the role of m^{6}A in aging regulation is warranted.

Conclusion

In this study, a total of thirteen m^{6}A writer genes in *G. max*, thirteen in *G. soja*, six in *P. vulgaris*, eight in *M. truncatula*, and six in *L. japonicus* were identified. The phylogenetic analysis divided

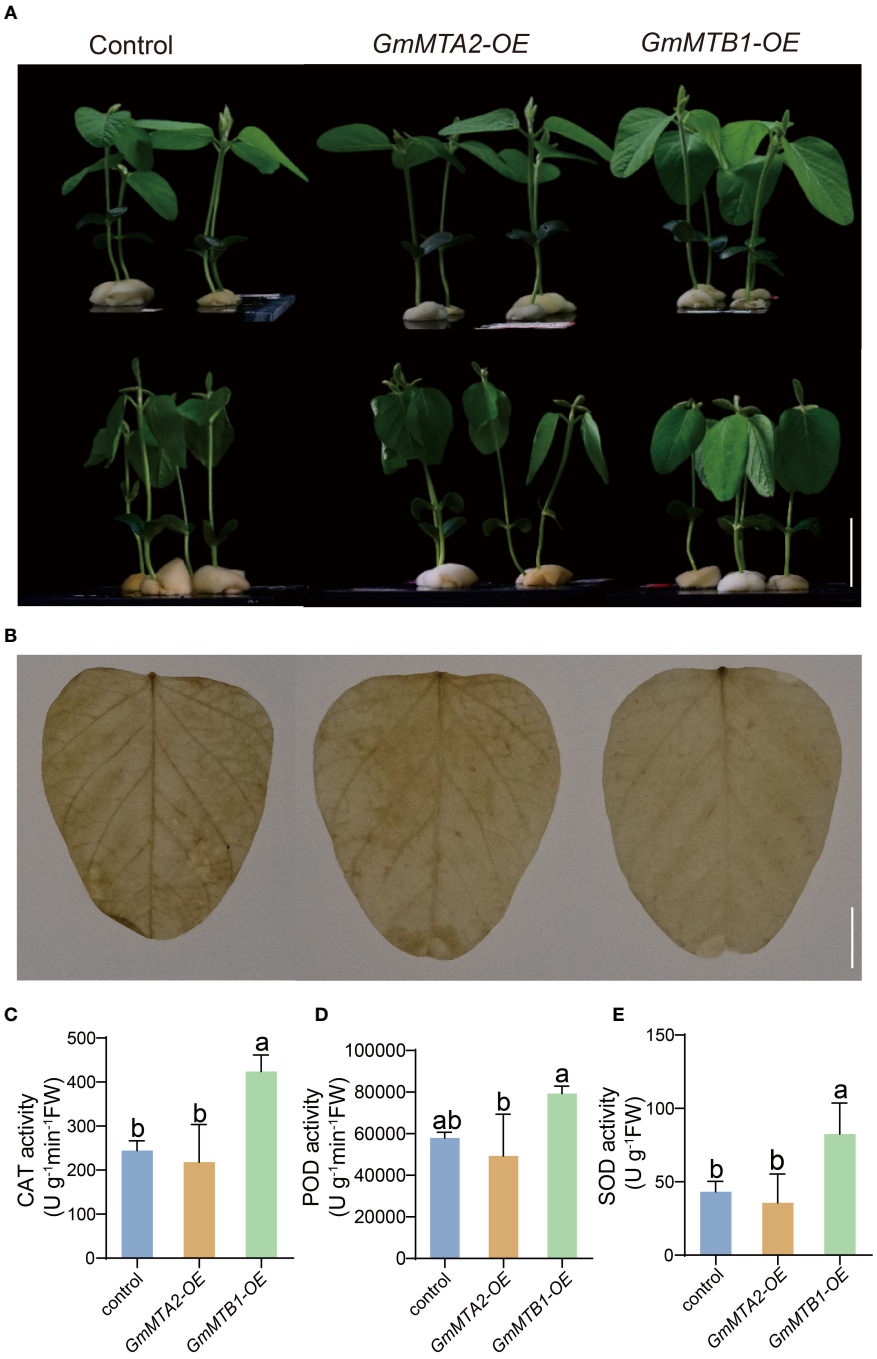


FIGURE 9
Leaf phenotypes of the empty vector-expressing control, *GmMTA2-OE*, and *GmMTB1-OE* plants under the $NaHCO_3$ treatment. **(A)** Plant documents before (top panel) and after (bottom panel) a 36-h $NaHCO_3$ treatment. Bar, 3 cm. **(B)** Diaminobenzidine staining of leaves after the $NaHCO_3$ treatment. Bar, 1 cm. **(C-E)** The catalase (CAT; C), peroxidase (POD; D), and superoxide dismutase (SOD; E) activities in the leaves after the $NaHCO_3$ treatment. Error bars represent SD (n = 3). Different letters indicate significant differences at $P < 0.05$.

these genes into four families based on their topological structure. In soybean, the collinearity analysis revealed members from each m^6A writer family originated from gene duplication. WoLF PSORT prediction coupled with subcellular localization analysis suggested that these m^6A writer genes all localized in nucleus. Furthermore, enzymatic analysis showed that both GmMTA1 and GmMTA2 possessed the methyltransferase activities toward adenosine on RNA. The cis-acting elements of 2000 bp promoter regions of all m^6A writer genes were investigated and the expression pattern of four MT-A70 family members upon abiotic stress treatment were determined. The results suggested that all members, especially *GmMTA2* and *GmMTB1*, were involved in the cellular response to various abiotic stress. Notably, soybean leaves overexpressing *GmMTB1* exhibited more resistant to the alkalinity, while

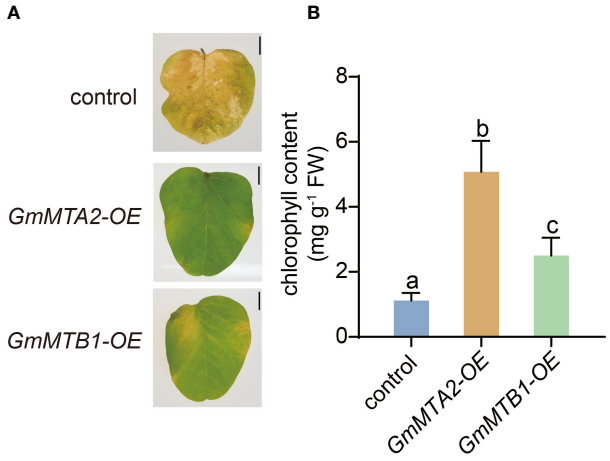


FIGURE 10
Leaf phenotypes of the control, *GmMTA2-OE*, and *GmMTB1-OE* plants upon darkness treatment. **(A)** The phenotypic changes of the leaves subjected to darkness for 10 days. Bar, 1 cm. **(B)** Chlorophyll contents of leaves from the dark-treated plants. Error bars represent SD (n = 3). Different letters indicate significant differences at $P < 0.05$.

overexpressing *GmMTA2* or *GmMTB1* both showed the highest tolerance to darkness treatment. These results will provide a basis for further exploring the biological functions of the m^6A writer genes from legume plants in growth regulation and stress responses.

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

PL: Data curation, Formal analysis, Methodology, Writing – original draft. HL: Writing – original draft, Data curation, Formal analysis. JZ: Data curation, Formal analysis, Methodology, Writing – original draft. TY: Data curation, Writing – original draft. SG: Data curation, Writing – original draft. LC: Data curation, Writing – original draft. TX: Data curation, Writing – original draft. AX: Data curation, Writing – original draft. XL: Formal analysis, Methodology, Writing – original draft. CZ: Formal analysis, Methodology, Writing – original draft. LG: Formal analysis, Methodology, Writing – original draft. MC: Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was financially supported by the National Natural Science Foundation of China (grant number 32371365), the Natural Science Foundation of Jiangsu Province, China (grant number BK20230101), the Project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions, and the start-up fund for advanced talents from Nanjing Agricultural University to MC.

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.2024.1446591/full#supplementary-material

References

Anderson, S. J., Kramer, M. C., Gosai, S. J., Yu, X., Vandivier, L. E., Nelson, A. D. L., et al. (2018). N6-methyladenosine inhibits local ribonucleolytic cleavage to stabilize mRNAs in Arabidopsis. *Cell Rep.* 25, 1146–1157. doi: 10.1016/j.celrep.2018.10.020

Baccolini, C., and Witte, C.-P. (2019). AMP and GMP catabolism in Arabidopsis converge on xanthosine, which is degraded by a nucleoside hydrolase heterocomplex. *Plant Cell* 31, 734–751. doi: 10.1105/tpc.18.00899

Cannon, S. B., Mitra, A., Baumgarten, A., Young, N. D., and May, G. (2004). The roles of segmental and tandem gene duplication in the evolution of large gene families in *Arabidopsis thaliana*. *BMC Plant Biol.* 4. doi: 10.1186/1471-2229-4-10

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, M., Herde, M., and Witte, C.-P. (2016). Of the nine cytidine deaminase like genes in *Arabidopsis thaliana* eight are pseudogenes and only one is required to maintain pyrimidine homeostasis *in vivo*. *Plant Physiol.* 171, 799–809. doi: 10.1104/pp.15.02031

Chen, M., Urs, M. J., Sánchez-González, I., Olayioye, M. A., Herde, M., and Witte, C.-P. (2018). m^{6}A RNA degradation products are catabolized by an evolutionarily conserved N^6-Methyl-AMP deaminase in plant and mammalian cells. *Plant Cell* 30, 1511–1522. doi: 10.1105/tpc.18.00236

Chen, M., and Witte, C.-P. (2020). A kinase and a glycosylase catabolize pseudouridine in the peroxisome to prevent toxic pseudouridine monophosphate accumulation. *Plant Cell* 32, 722–739. doi: 10.1105/tpc.19.00639

Cheng, C., Liu, Y., Liu, X., An, J., Jiang, L., and Yu, B. (2019). Recretohalophyte Tamarix *TrSOS1* confers higher salt tolerance to transgenic plants and yeast than glycophyte soybean *GmSOS1*. *Environ. Exp. Bot.* 165, 196–207. doi: 10.1016/j.envexpbot.2019.06.006

Delaunay, S., Helm, M., and Frye, M. (2023). RNA modifications in physiology and disease: towards clinical applications. *Nat. Rev. Genet.* 25, 104–122. doi: 10.1038/s41576-023-00645-2

Feng, N., Yu, M., Li, Y., Jin, D., and Zheng, D. (2021). Prohexadione-calcium alleviates saline-alkali stress in soybean seedlings by improving the photosynthesis and up-regulating antioxidant defense. *ECOTOX Environ. SAFE* 220, 112369. doi: 10.1016/j.ecoenv.2021.112369

Frye, M., Harada, B. ,. T., Behm, M., and He, C. (2018). RNA Modifications modulate gene expression during development. *Science* 361, 1346–1349. doi: 10.1126/science.aau1646

Frye, M., Jaffrey, S. R., Pan, T., Rechavi, G., and Suzuki, T. (2016). RNA modifications: what have we learned and where are we headed? *Nat. Rev. Genet.* 17, 365–372. doi: 10.1038/nrg.2016.47

Fu, Y., Guo, C., Wu, H., and Chen, C. (2017). Arginine decarboxylase ADC2 enhances salt tolerance through increasing ROS scavenging enzyme activity in *Arabidopsis thaliana*. *Plant Growth Regul.* 83, 253–263. doi: 10.1007/s10725-017-0293-0

Gao, S., Sun, Y., Chen, X., Zhu, C., Liu, X., Wang, W., et al. (2023). Pyrimidine catabolism is required to prevent the accumulation of 5-methyluridine in RNA. *Nucleic Acids Res.* 51, 7451–7464. doi: 10.1093/nar/gkad529

Guo, Y., Ren, G., Zhang, K., Li, Z., Miao, Y., and Guo, H. (2021). Leaf senescence: progression, regulation, and application. *Mol. Hortic.* 1, 5. doi: 10.1186/s43897-021-00006-9

Guo, T., Yang, Z., Bao, R., Fu, X., Wang, N., Liu, C., et al. (2023). The m^{6}A reader MhYTP2 regulates the stability of its target mRNAs contributing to low nitrogen tolerance in apple (*Malus domestica*). *Hortic* 10, uhad094. doi: 10.1093/hr/uhad094

Han, B., Wei, S., Li, F., Zhang, J., Li, Z., and Gao, X. (2021). Decoding m^{6}A mRNA methylation by reader proteins in cancer. *Cancer Lett.* 518, 256–265. doi: 10.1016/j.canlet.2021.07.047

Hou, N., Li, C., He, J., Liu, Y., Yu, S., Malnoy, M., et al. (2022). MdMTA-mediated m^{6}A modification enhances drought tolerance by promoting mRNA stability and translation efficiency of genes involved in lignin deposition and oxidative stress. *New Phytol.* 234, 1294–1314. doi: 10.1111/nph.18069

Hu, J., Cai, J., Park, S. J., Lee, K., Li, Y., Chen, Y., et al. (2021). N6-Methyladenosine mRNA methylation is important for salt stress tolerance in Arabidopsis. *Plant J.* 106, 1759–1775. doi: 10.1111/tpj.15270

Ji, X., Liu, G., Liu, Y., Zheng, L., Nie, X., and Wang, Y. (2013). The bZIP protein from Tamarix hispida, ThbZIP1, is ACGT elements binding factor that enhances abiotic stress signaling in transgenic Arabidopsis. *BMC Plant Biol.* 13, 151. doi: 10.1186/1471-2229-13-151

Jia, G., Fu, Y., and He, C. (2013). Reversible RNA adenosine methylation in biological regulation. *Trends Genet.* 29, 108–115. doi: 10.1016/j.tig.2012.11.003

Katiyar, A., Smita, S., Lenka, S. K., Rajwanshi, R., Chinnusamy, V., and Bansal, K. C. (2012). Genome-wide classification and expression analysis of MYB transcription factor families in rice and Arabidopsis. *BMC Genom* 13, 544. doi: 10.1186/1471-2164-13-544

Kim, E., Hwang, S., and Lee, I. (2017). SoyNet: a database of co-functional networks for soybean *Glycine max*. *Nucleic Acids Res.* 45, D1082–D1089. doi: 10.1093/nar/gkw704

Kost, B., Spielhofer, P., and Chua, N. H. (2008). A GFP-mouse talin fusion protein labels plant actin filaments *in vivo* and visualizes the actin cytoskeleton in growing pollen tubes. *Plant J.* 16, 393–401. doi: 10.1046/j.1365-313x.1998.00304.x

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, 2947–2948. doi: 10.1093/bioinformatics/btm404

Li, Z., Lyu, X., Li, H., Tu, Q., Zhao, T., Liu, J., et al. (2024). The mechanism of low blue light-induced leaf senescence mediated by GmCRY1s in soybean. *Nat. Commun.* 15, 798. doi: 10.1038/s41467-024-45086-5

Lim, C., Kang, K., Shim, Y., Yoo, S. C., and Paek, N. C. (2022). Inactivating transcription factor OsWRKY5 enhances drought tolerance through abscisic acid signaling pathways. *Plant Physiol.* 188, 1900–1916. doi: 10.1093/plphys/kiab492

Liu, Y., Zhang, Y., Liu, X., Shen, Y., Tian, D., Yang, X., et al. (2023). SoyOmics: A deeply integrated database on soybean multi-omics. *Mol. Plant* 16, 794–797. doi: 10.1016/j.molp.2023.03.011

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

Lu, L., Zhang, Y., He, Q., Qi, Z., Zhang, G., Xu, W., et al. (2020). MTA, an RNA m^{6}A Methyltransferase, enhances drought tolerance by regulating the development of trichomes and roots in Poplar. *Int. J. Mol. Sci.* 21, 2462. doi: 10.3390/ijms21072462

Martin, K., Kopperud, K., Chakrabarty, R., Banerjee, R., Brooks, R., and Goodin, M. M. (2009). Transient expression in Nicotiana benthamiana fluorescent marker lines provides enhanced definition of protein localization, movement and interactions in planta. *Plant J.* 59, 150–162. doi: 10.1111/j.1365-313X.2009.03850.x

Phukan, U. J., Jeena, G. S., and Shukla, R. K. (2016). WRKY transcription factors: molecular regulation and stress responses in plants. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.00760

Pi, B., Liu, X., Huang, Q., Zhang, T., and Yu, B. (2023). Comparative transcriptomic analysis of *Glycine soja* and *G. max* and functional identification of *GsCNGC20-d* interacted with *GsCDPK29* under salt stress. *Environ. Exp. Bot.* 206, 105185. doi: 10.1016/j.envexpbot.2022.105185

Qi, D. H., and Lee, C. F. (2014). Influence of soybean biodiesel content on basic properties of biodiesel-diesel blends. *J. Taiwan Inst. Chem. Eng.* 45, 504–507. doi: 10.1016/j.jtice.2013.06.021

Qu, Y., Guan, R., Yu, L., Berkowitz, O., David, R., Whelan, J., et al. (2022). Enhanced reactive oxygen detoxification occurs in salt-stressed soybean roots expressing GmSALT3. *Physiol. Plant* 174, e13709. doi: 10.1111/ppl.13709

Roundtree, I. A., Evans, M. E., Pan, T., and He, C. (2017). Dynamic RNA modifications in gene expression regulation. *Cell* 169, 1187–1200. doi: 10.1016/j.cell.2017.05.045

Růžička, K., Zhang, M., Campilho, A., Bodi, Z., Kashif, M., Saleh, M., et al. (2017). Identification of factors required for m^{6}A mRNA methylation in Arabidopsis reveals a role for the conserved E3 ubiquitin ligase HAKAI. *New Phytol.* 215, 157–172. doi: 10.1111/nph.14586

Sakuraba, Y., Jeong, J., Kang, M.-Y., Kim, J., Paek, N.-C., and Choi, G. (2014). Phytochrome-interacting transcription factors PIF4 and PIF5 induce leaf senescence in Arabidopsis. *Nat. Commun.* 5, 4636. doi: 10.1038/ncomms5636

Sheikh, A. H., Tabassum, N., Rawat, A., Almeida, T. M., Nawaz, K., and Hirt, H. (2024). m^{6}A RNA methylation counteracts dark-induced leaf senescence in Arabidopsis. *Plant Physiol.* 194, 2663–2678. doi: 10.1093/plphys/kiad660

Shen, L., Liang, Z., Gu, X., Chen, Y., Teo, Z. W. N., Hou, X., et al. (2016). N6-methyladenosine RNA modification regulates shoot stem cell fate in Arabidopsis. *Dev. Cell* 38, 186–200. doi: 10.1016/j.devcel.2016.06.008

Shi, H., Wei, J., and He, C. (2019). Where, When, and How: context-dependent functions of RNA methylation writers, readers, and erasers. *Mol. Cell* 74, 640–650. doi: 10.1016/j.molcel.2019.04.025

Song, Y., Yang, C., Gao, S., Zhang, W., Li, L., and Kuai, B. (2014). Age-triggered and dark-induced leaf senescence require the bHLH transcription factors PIF3, 4, and 5. *Mol. Plant* 7, 1776–1787. doi: 10.1093/mp/ssu109

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

van Zelm, E., Zhang, Y., and Testerink, C. (2020). Salt tolerance mechanisms of plants. *Annu. Rev. Plant Biol.* 71, 403–433. doi: 10.1146/annurev-arplant-050718-100005

Vespa, L., Vachon, G., Berger, F., Perazza, D., Faure, J.-D., and Herzog, M. (2004). The immunophilin-interacting protein AtFIP37 from Arabidopsis is essential for plant development and is involved in trichome endoreduplication. *Plant Physiol.* 134, 1283–1292. doi: 10.1104/pp.103.028050

Wang, X., Feng, J., Xue, Y., Guan, Z., Zhang, D., Liu, Z., et al. (2016b). Structural basis of N6-adenosine methylation by the METTL3–METTL14 complex. *Nature* 534, 575–578. doi: 10.1038/nature18298

Wang, T., Guo, J., Peng, Y., Lyu, X., Liu, B., Sun, S., et al. (2023a). Light-induced mobile factors from shoots regulaterhizobium-triggered soybean root nodulation. *Science* 374, 65–71. doi: 10.1126/science.abh2890

Wang, W., Liu, H., Wang, F., Liu, X., Sun, Y., Zhao, J., et al. (2023b). N4-acetylation of cytidine in mRNA plays essential roles in plants. *Plant Cell* 35, 3739–3756. doi: 10.1093/plcell/koad189

Wang, F., Liu, J., Zhou, L., Pan, G., Li, Z., Zaidi, S.-H.-R., et al. (2016a). Senescence-specific change in ROS scavenging enzyme activities and regulation of various SOD isozymes to ROS levels in psf mutant rice leaves. *Plant Physiol. Biochem.* 109, 248–261. doi: 10.1016/j.plaphy.2016.10.005

Woo, H. R., Kim, H. J., Lim, P. O., and Nam, H. G. (2019). Leaf senescence: systems and dynamics aspects. *Annu. Rev. Plant Biol.* 70, 347–376. doi: 10.1146/annurev-arplant-050718-095859

Wu, Z., Shi, Y., Lu, M., Song, M., Yu, Z., Wang, J., et al. (2020). METTL3 counteracts premature aging via m^6A-dependent stabilization of MIS12 mRNA. *Nucleic Acids Res.* 48, 11083–11096. doi: 10.1093/nar/gkaa816

Xie, J., Chen, Y., Cai, G., Cai, R., Hu, Z., and Wang, H. (2023). Tree Visualization By One Table (tvBOT): a web application for visualizing, modifying and annotating phylogenetic trees. *Nucleic Acids Res.* 51, W587–W592. doi: 10.1093/nar/gkad359

Xing, K., Liu, Z., Liu, L., Zhang, J., Qanmber, G., Wang, Y., et al. (2023). N6-methyladenosine mRNA modification regulates transcripts stability associated with cotton fiber elongation. *Plant J.* 115, 967–985. doi: 10.1111/tpj.16274

Yang, L., Han, Y., Wu, D., Yong, W., Liu, M., Wang, S., et al. (2017). Salt and cadmium stress tolerance caused by overexpression of the *Glycine Max* Na+/H+ Antiporter (*GmNHX1*) gene in duckweed (*Lemna turionifera 5511*). *Aquat. Toxicol.* 192, 127–135. doi: 10.1016/j.aquatox.2017.08.010

Zheng, H., Sun, X., Li, J., Song, Y., Song, J., Wang, F., et al. (2021). Analysis of N6-methyladenosine reveals a new important mechanism regulating the salt tolerance of sweet sorghum. *Plant Sci.* 304, 110801. doi: 10.1016/j.plantsci.2020.110801

Zhong, S., Li, H., Bodi, Z., Button, J., Vespa, L., Herzog, M., et al. (2008). MTA is an Arabidopsis messenger RNA adenosine methylase and interacts with a homolog of a sex-specific splicing factor. *Plant Cell* 20, 1278–1288. doi: 10.1105/tpc.108.058883

Zhou, L., Gao, G., Tang, R., Wang, W., Wang, Y., Tian, S., et al. (2022). m^6A-mediated regulation of crop development and stress responses. *Plant Biotechnol.* 20, 1447–1455. doi: 10.1111/pbi.13792

Zhu, M., Hu, Z., Zhou, S., Wang, L., Dong, T., Pan, Y., et al. (2014). Molecular characterization of six tissue-specific or stress-inducible genes of NAC transcription factor family in Tomato (*Solanum lycopersicum*). *J. Plant Growth Regul.* 33, 730–744. doi: 10.1007/s00344-014-9420-6

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 26 September 2024
DOI 10.3389/fpls.2024.1487092

OPEN ACCESS

EDITED BY
Huatao Chen,
Jiangsu Academy of Agricultural Sciences (JAAS), China

REVIEWED BY
Chengfu Su,
Qingdao Agricultural University, China
Li Song,
Yangzhou University, China

*CORRESPONDENCE
Huanli Li
lihuanli198906@163.com

RECEIVED 27 August 2024
ACCEPTED 09 September 2024
PUBLISHED 26 September 2024

CITATION
Li H, Zhang X, Yang Q, Shangguan X and Ma Y (2024) Genome-wide identification and tissue expression pattern analysis of *TPS* gene family in soybean (*Glycine max*).
Front. Plant Sci. 15:1487092.
doi: 10.3389/fpls.2024.1487092

Genome-wide identification and tissue expression pattern analysis of *TPS* gene family in soybean (*Glycine max*)

Huanli Li*, Xiaoling Zhang, Qinli Yang, Xiaoxia Shangguan and Yanbin Ma

Cotton Research Institute of Shanxi Agricultural University, Yuncheng, China

The terpene synthase (TPS) plays a pivotal roles in plant growth, development, and enhancing resilience against environmental stresses. Despite this, the bioinformatics analysis of the *TPS* family gene in soybean (*Glycine max*) is lacking. In this study, we investigated 36 GmTPS members in soybean, exhibiting a diverse range of protein lengths, spanning from 144 to 835 amino acids. A phylogenetic tree was constructed from these *GmTPS* genes revealed a classification into five distinct subgroups: Group1, Group2, Group3, Group4 and Group5. Notably, within each subgroup, we identified the motifs of GmTPS proteins were similar, although variations existed among different subfamilies. Gene duplication events analysis demonstrated that *TPS* genes expand differently in *G. max*, *A. thaliana* and *O. sativa*. Among, both tandem duplication and Whole genome duplication contributive to the expansion of *TPS* genes in *G. max*, and Whole genome duplication played a major role. Moreover, the cis-element analysis suggested that *TPS* is related to hormone signals, plant growth and development and environmental stress. Yeast two-hybrid (Y2H) assay results indicated TPS protein may form heterodimer to function, or may form complex with P450 proteins to function. RNA-seq results revealed a higher expression of most *GmTPS* genes in flowers, suggesting their potential contribution to flower development. Collectively, these findings offer a provide a holistic knowledge of the TPS gene family in soybean and will facilitate further characterization of *TPSs* effectively.

KEYWORDS

***Glycine max*, TPS gene family, phylogenetic analysis, duplicated events, RNA-seq**

1 Introduction

Terpenoids, also referred to as isoprenoids, are abundant natural products, and more than 80,000 terpenoids and their derivatives have been found so far, widely existing in plants, fungi, bacteria and insects (Realdon, 1960). TPS proteins are widely found in *algae*, bryophytes, ferns, monocotyledons and dicotyledons. Kaul et al. (2000) cloned the first TPS

enzymes coding gene *AtTPS1* in *A. thaliana*, with the development of modern sequencing technologies, and more and more *TPS* genes were identified in plant genomes. For instance, there are 33 *TPS* genes in *Arabidopsis* (Aubourg et al., 2002; Yang et al., 2012), 53 in rice (Yang et al., 2012), 12 in *populus* (Yang et al., 2012), 8 in *potato* (Xu et al., 2017), 9 in *B. distachyon* (Wang et al., 2019), 34 in *d. officinale* (Yu et al., 2020), 80 in *camellia* (Zhou et al., 2020), 26 in *aloes* (Li et al., 2021), 58 in *l. chinense* (Cao et al., 2023) and 16 in *A. hypogaea* (Zhong et al., 2024). Notably, many TPS genes are also found in bacteria (Jia et al., 2019).

Previous research has categorized TPS proteins into 7 subfamilies: TPS-a–TPS-h. Specifically, TPS-a, TPS-b and TPS-g are present in angiosperms; TPS-c, closely related to TPS-e/f, is related in diterpenoid synthase production and is found in gymnosperms. Meanwhile, TPS-e/f is present in vascular plants (Newman and Chappell, 1999; Chen et al., 2011).

The regulation of *TPS* gene expression is influenced by various hormonal and environmental stresses. Such as, MeJA treatments upregulate most *CsTPS* genes expression (Zhou et al., 2020), while osmotic stress and heat stress induce *TPS* genes upregulation in roses (Yan et al., 2022). Transgenic studies show that *TPS* overexpression can enhance stress tolerance in crops like rice and *Arabidopsis*. For instance, *OsTPS1* overexpression improves rice low temperature tolerance (Ge et al., 2008) and *AhTPS9* overexpression enhances *Arabidopsis* cold tolerance (Zhong et al., 2024). Similarly, *ScTPS1* overexpression in potatoes improves drought tolerance (Yeo et al., 2000). Moreover, *OsTPS46* confers natural resistance to bird cherry-oat aphid (Sun et al., 2017), while *OsTPS24* showed no significant inhibitory activity against Magnaporthe oryzae (Yoshitomi et al., 2016). In soybeans, *GmAFS* have defensive effects against nematodes and insects (Lin et al., 2017).

Despite the importance of *TPS* genes in stress resistance, their functions in soybean remain largely unexplored. Here, we carried out a bioinformatics analysis of the *TPS* gene family in soybean, examining phylogenetic relationships, gene structures, duplication events, gene collinearity, and protein interaction networks. The tissue expression of *GmTPS* in the 6 tissues of root, young leaf, pod shell, flower, seed and nodule unravel their key regulational roles during soybean development. This study offers valuable insights and theoretical support for understanding the roles of *GmTPS* genes in soybean stress resistance.

2 Materials and methods

2.1 Data sources and identification of TPSs in soybean

The genome data were downloaded from the Soybean database (https://www.soybase.org/dlpages/). The AtTPSs and OsTPSs protein sequences were downloaded from TAIR and RGAP, respectively. The hidden Markov Model (HMM) file of PF01397, PF03936 and PF19086 were downloaded from InterPro database (Paysan et al., 2023). Utilizing HMMER 3.0, we screened for TPS proteins within soybean (E-value <= 1e-5, similarity > 50%) (Mistry et al., 2013). Additionally, we employed the BLASTP method (Camacho et al., 2009) to search GmTPS protein sequences using AtTPS and OsTPS proteins as references (E-value <= 1e-5, similarity > 50%). Subsequently, the identified candidate TPS protein sequences underwent domain verification, adopting the analysis approach outlined by Xu et al. (2017). We filtered the longest transcript using the R package seqfinder (https://github.com/yueliu1115/seqfinder).

2.2 Evolutionary trees are constructed of GmTPS, AtTPS and OsTPS

We employed the Muscle software (Edgar, 2004) for multiple sequence alignment of GmTPS, AtTPS and OsTPS proteins, and constructed a phylogenetic tree using IQ-TREE (Nguyen et al., 2015). The tree was visualized using the R package ggtree (Yu et al., 2017).

2.3 The cis-elements analysis of GmTPS genes

The 2 kb promoter region sequences upstream of the *GmTPS* gene were extracted using a Python program and submitted to the PlantCare (https://bioinformatics.psb.ugent.be/webtools/plantcare/html/) database for cis-element prediction (Lescot et al., 2002). All results were visualized in R software.

2.4 Analysis of chromosome distribution, gene duplication events, and selection pressure

The chromosomal distribution of *GmTPS* genes was derived from the soybean genome annotation information. MCScanX software (Wang et al., 2012) was utilized for gene duplication and colinearity analysis, identifying duplication types such as tandem (TD), and whole genome (WGD). For interspecies collinearity analysis and visualization, JCVI software (Tang et al., 2008) was employed. JCVI software was used for interspecies collinearity analysis and visualization (Tang et al., 2008). ClustalW software was used to align the protein sequences and CDS sequences of *TPS* genes with gene duplication (Thompson et al., 2003). KaKs_Calculator software was used to calculate the synonymous substitution rate (synonymous, Ks), nonsynonymous substitution rate (nonsynonymous, Ka) and evolutionary ratio (Ka/Ks) between *TPS* genes duplicate gene pairs (Zhang, 2022).

2.5 TPS protein interaction network analysis

The GmTPS proteins interaction network were predicted based on the AraNet2 database (Lee et al., 2015).

2.6 Tissue expression pattern analysis of GmTPS genes using RNA-seq

The transcriptome data of soybean under different tissues and development stages from the Soybean database (https://www.soybase.org/dlpages/). The expression data were visualized using the R package Pheatmap (Kolde and Kolde, 2015).

2.7 Yeast two-hybrid assays

The CDS sequence of the *Glyma.07G192800* and *Glyma.15G263300* were cloned into the pGBKT7 vector (BD-TPS); the CDS sequence of the Glyma.12G140600, Glyma.09G029400, Glyma.20G074400 and Glyma.01G153300 were cloned into the pGADT7 vector (AD-TPSs or AD-p450s). Yeast transformants with empty pGBKT7 and AD-TPSs or AD-P450s; yeast transformants with empty pGADT7 and BD-TPSs were used as the negative control. The positive control: AD-T + BD-53. Yeast transformants with AD-TPSs, BD-TPSs and AD-p450s were used to identify the TPSs interact with other TPS proteins or P450 proteins.

3 Results

3.1 Identification of TPS members in soybean

Here, a total of 36 TPS members in soybean were identified by HMMER and BLASTP methods (Table 1). The length of the GmTPSs protein sequence ranged from 144 (scaffold_311) to 835 (Glyma.08G163900.1); the molecular weight ranged from 16.3 (scaffold_311) to 95.6 KDa (Glyma.13G183600.1); the PI ranged from 4.27 (Glyma.13G304700.1) to 8.45 (Glyma.20G248300.1) (Table 1). It's worth noting that 86% of GmTPS have isoelectric points less than 7 (Table 1), it is suggest that most *GmTPS* genes may are acidic proteins.

3.2 Phylogenetic analysis of TPSs

To deeper understanding the evolutionary dynamics of GmTPSs, we constructed a phylogenetic tree that comprehensively encompasses 36 GmTPSs, along with 33 AtTPSs and 53 OsTPSs. The *TPS*s members could be grouped into Group1, Group2, Group3, Group4 and Group5 (Figure 1). Group5 contained the largest number of 35 *TPS*s, while Group3 contained the smallest number of 10 *TPS*s. Group1, Group2, and Group4 contained the number of 19, 27 and 32 *TPS*s, respectively. Interestingly, no *GmTPS*s and At*TPS*s member was found in the subgroup of Group5, and Group1 and Group4 contain only GmTPS and AtTPS members (Figure 1). These observations provide insights into the evolution of the *TPS* gene family.

3.3 Gene structure and motifs analysis of GmTPS

To comprehend the diversity of *GmTPS* genes, we analyzed the gene structures, and conserved domain and conserved motif in GmTPS, the results are presented in Figure 2. Most of the GmTPS members contained two conserved domains, the Terpene_synth domain at the N-terminal and the Terpene_synth_C domain at the C-terminal (Figure 2). It is worth noting that Glyma.U032900 and Glyma.07G192800 consist only of the Terpene_synth_C domain; Glyma.19G156800 consist only of the Terpene_synth domain. Interestingly, Glyma.06G291800 contain two Terpene_synth_C domains (Figure 2). Except for the Terpene_synth_C domains, Glyma.12G179500 also contains Terpene_syn_C_2 domain.

On the other hand, in the same subgroups, we found the conserved motifs of GmTPS were similar, although variations existed among different subfamilies (Figure 2). For example, Group4 Subgroup members Glyma.12G138600, Glyma.12G140600, Glyma.12G138800, Glyma.12G216200 and Glyma.13G285200 contain conserved motif 1, motif 2, motif 3, motif 5, motif 6, motif 7, motif 10, motif 11, motif 14 and motif 15; Glyma.12G197500 and Glyma.12G197400 contain conserved motif 1, motif 2, motif 3, motif 4, motif 6, motif 7, motif 9, motif 10 motif 11 and motif 14 (Figure 2). These diverse motifs reflect the functional diversity of GmTPS proteins. On the other hand, we found that most *TBS* genes contain multiple introns, except for *Glyma.U032900* and *Glyma.13G304800*, which contain only one intron (Figure 2). The variations of motifs and gene structure may contribute to the diverse biological functions of *GmTPSs*.

3.4 Duplication events analysis of GmTPS genes

According to GFF files, we analyzed the gene distribution of 36 *GmTPS*. The 36 *GmTPS* genes were distributed on 10 chromosomes, while no *GmTPS* genes were distributed on chromosomes Chr1, Chr2, Chr4, Chr5, Chr10, Chr11, Chr14, Chr16, Chr17 and Chr18 (Figure 3). Notably, the most *GmTPS* genes were distributed at chr12 and chr13, with 10 and 19 respectively, forming gene clusters. Both Chr9 and Chr15 contain one *GmTPS* members, while both Chr3, Chr6, Chr8, Chr19 and Chr20 contain two *GmTPS* members (Figure 3).

TD and WGD drive the expansion of the gene family (Freeling, 2009; Panchy et al., 2016). Therefore, we explored the duplication events of *TPS* genes in soybean, *Arabidopsis* and rice. In this study, 16 WGD gene pairs and 18 TD gene pairs were confirmed in soybean, *Arabidopsis* and rice (Figures 3A–C; Supplementary Table 1). Overall, in soybean, *Arabidopsis* and rice, 10 (27.78%), 18 (60%) and 16 (30.19%) *TPS* genes were confirmed to be TD, and 23 (63.89%), 2 (6%) and 0 (0%) *TPS* genes were found to be WGD, respectively (Figures 3A–C; Supplementary Table 1). These data show that both TD and WGD contributive to the expansion of *TPS* genes in soybean, and WGD played a major role. However, in

10.3389/fpls.2024.1487092

TABLE 1 Physical and chemical property analysis of TPS family genes in soybean (*Glycine max*).

Gene ID	Chr	Start	End	Amino acid length	MW	pI	Hydrophobicity
Glyma.03G154400.1	Gm03	36951469	36958622	769	87865.15	5.68	−0.28
Glyma.03G154700.1	Gm03	36991751	37001866	816	93579.29	6.70	−0.36
Glyma.06G291800.1	Gm06	48044193	48047197	324	37696.52	4.81	−0.24
Glyma.06G302200.1	Gm06	49127197	49132358	598	68742.52	6.68	−0.31
Glyma.07G187600.1	Gm07	35502526	35506056	574	65884.17	5.58	−0.26
Glyma.07G187700.1	Gm07	35531688	35536600	589	67601.58	6.05	−0.24
Glyma.07G192800.1	Gm07	36063173	36066134	380	44365.98	8.02	−0.39
Glyma.08G061600.1	Gm08	4751072	4754299	292	33490.34	5.43	−0.18
Glyma.08G163900.1	Gm08	12901875	12908991	835	95150.30	6.60	−0.18
Glyma.09G122500.1	Gm09	29418942	29427117	603	69487.15	6.36	−0.41
Glyma.10G297200.1	Gm10	51407332	51408418	225	25981.68	6.42	−0.32
Glyma.12G101700.1	Gm12	8981010	8985722	377	43496.38	5.65	−0.03
Glyma.12G102000.1	Gm12	9043334	9048807	603	69766.91	7.08	−0.32
Glyma.12G138100.1	Gm12	16508959	16517078	531	61954.20	5.62	−0.27
Glyma.12G138600.1	Gm12	16685178	16690007	554	64037.47	5.91	−0.19
Glyma.12G138800.1	Gm12	16733572	16741825	501	58069.48	6.38	−0.34
Glyma.12G140600.1	Gm12	17392329	17399822	561	65363.99	5.37	−0.24
Glyma.12G179500.1	Gm12	33977724	33980896	421	48739.09	6.65	−0.30
Glyma.12G197400.1	Gm12	35854107	35858237	569	65013.70	6.91	−0.26
Glyma.12G197500.1	Gm12	35869161	35874097	585	67079.78	5.91	−0.27
Glyma.12G216200.1	Gm12	37539186	37543933	565	65213.59	5.64	−0.27
Glyma.13G183600.1	Gm13	29714220	29722406	832	95610.00	7.08	−0.27
Glyma.13G250400.1	Gm13	35790066	35794185	535	62338.13	5.87	−0.29
Glyma.13G285100.1	Gm13	38608137	38609427	256	29739.02	6.18	−0.31
Glyma.13G285200.1	Gm13	38611052	38614182	566	64535.43	5.56	−0.24
Glyma.13G304500.1	Gm13	40139730	40145370	580	66250.99	6.46	−0.26
Glyma.13G304600.1	Gm13	40155002	40157401	310	36171.91	6.19	−0.52
Glyma.13G304700.1	Gm13	40159803	40161491	233	27231.76	4.27	−0.33
Glyma.13G304800.1	Gm13	40161767	40162729	176	20013.88	6.51	−0.36
Glyma.13G321100.1	Gm13	41540416	41543961	569	65844.55	6.51	−0.26
Glyma.15G263300.1	Gm15	49639950	49645729	424	48964.87	5.06	−0.25
Glyma.19G156800.1	Gm19	41726563	41730894	291	33626.12	7.13	−0.35
Glyma.19G157000.2	Gm19	41763615	41774896	817	93957.76	6.64	−0.36
Glyma.20G074400.1	Gm20	26716996	26720775	607	70199.87	5.94	−0.46
Glyma.20G248300.1	Gm20	47752687	47755022	297	34026.17	8.45	−0.32
Glyma.U032900.1	scaffold_311	1109	2723	144	16302.60	6.09	−0.27

10.3389/fpls.2024.1487092

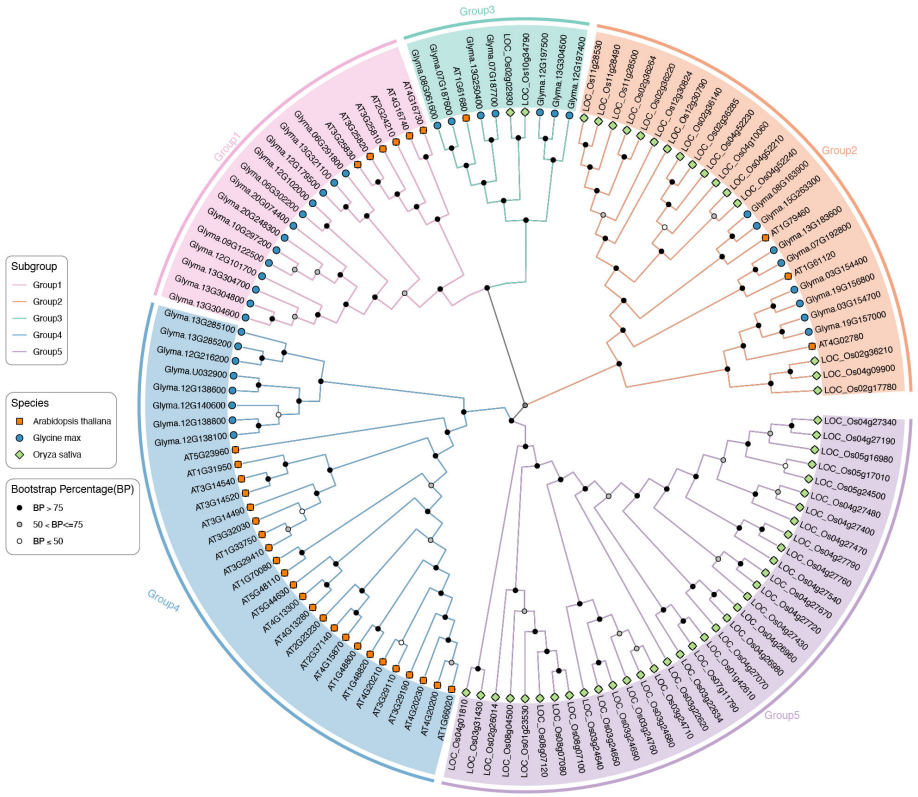


FIGURE 1
Phylogenetic tree of 36 GmTPSs, 33 AtTPSs and 53 ScTPSs. The evolutionary tree was constructed by the maximum likelihood method. Pink, orange, sky blue, blue and purple represent the subfamily of Group1, Group2, Group3, Group4 and Group5, respectively.

Arabidopsis, TD and WGD both promoted the expansion of *TPS* genes, and TD plays a leading role. Interestingly, in rice, only TD replication events were found. These data suggest that *TPS* genes expand differently in soybean, *Arabidopsis* and rice.

3.5 Collinearity analysis of *GmTPS* genes

To deeper investigate the homology of the *TPS* gene family in *G. max*, we conducted a comparative analysis of *TPS* gene collinearity between *G. max* and two model organisms, *Arabidopsis* and rice. Our findings revealed that 2 *AtTPS* and 1 *OsTPS* were homologous gene pairs with *GmTPS* (Figure 4). To gain insights into the evolutionary pressures of *TBS* genes in soybean, *Arabidopsis* and rice, we employed DnaSP software to calculate Ka/Ks ratios. Here, we found that the Ka/Ks value of all *TPS* duplication gene pairs is less than 1 in soybean, *Arabidopsis* and rice (Supplementary Table 1). These results suggest that *GmTPS*, *AtTPS* and *OsTPS* genes were under purifying selection.

3.6 Cis-elements analysis of GmTPS

To unravel the may regulatory mechanisms of *GmTPS genes*, we analyzed their 2k promoter regions, uncovering a diverse array of 21 cis-elements (Figure 5). These elements encompass various functional categories, including light response-related cis-elements (CAAT-box, Box-4, e.g); plant growth and developmental; phytohormone (ABRE, CGTCA-motif, TGACG-motif and TCA-element) and stress response related cis-elements (ARE, as-1, WUN-motif, MBS and TC-rich repeats) (Figure 5). Interestingly, Group4 *GmTPS* genes contain ABRE, CGTCA-motif and TGACG-motif cis-elements (Figure 5), hinting that *GmTPS* genes may be involved in ABA and JA signaling pathways. In addition, *Glyma.13G285100* contains CGTCA-motif, TGACG-motif and TCA-element, suggesting that it may antagonistically participate in SA and JA signaling pathways (Figure 5). It's worth noting that *Glyma.19G157000* contains a large number of light, plant growth and developmental and stress response cis-elements, while no hormone response cis-elements are found. Furthermore, our analysis revealed that *Glyma.07G192800*, *Glyma.12G138100*, *Glyma.15G263300*, and *Glyma.09G122500* contain varying numbers of MBS cis-elements, indicative of potential roles in drought signaling pathways (Figure 5). The above data indicates that *GmTPS* may have complex regulatory functions.

3.7 Interaction network of GmTPS proteins

To gain deeper insights into the functional roles and regulatory intricacies within the *GmTPS* gene family, we leveraged the

10.3389/fpls.2024.1487092

FIGURE 2
The domains, motifs and gene structure of GmTPSs were analyzed. **(A)** Phylogenetic tree, **(B)** The domains were predicted by NCBI-CDD. **(C)** The motifs of GmTPSs were predicted by MEME. **(D)** Gene structure of *GmTPS*.

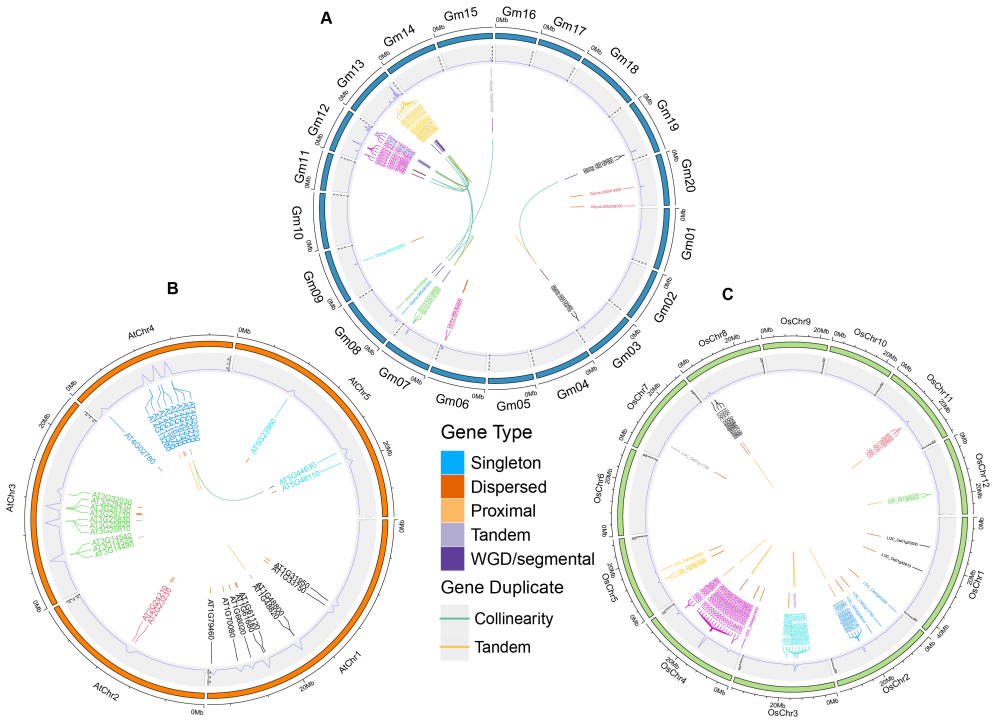


FIGURE 3
The chromosome location and duplicated gene pair of *TPS* genes in three species including **(A)** soybean, **(B)** *Arabidopsis* and **(C)** rice. The duplicate gene types were displayed in different color. WGD and TD events are shown in orange and blue, respectively.

10.3389/fpls.2024.1487092

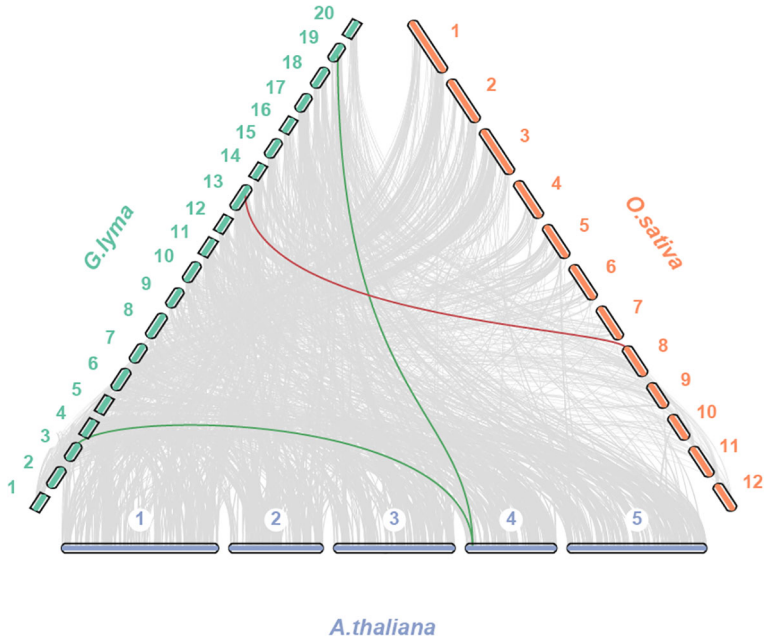


FIGURE 4
Syntenic analysis of *GmTPS* genes between and *Arabidopsis* and rice. The collinear blocks and *TPS* homologous genes pairs were shown by gray and red lines, respectively.

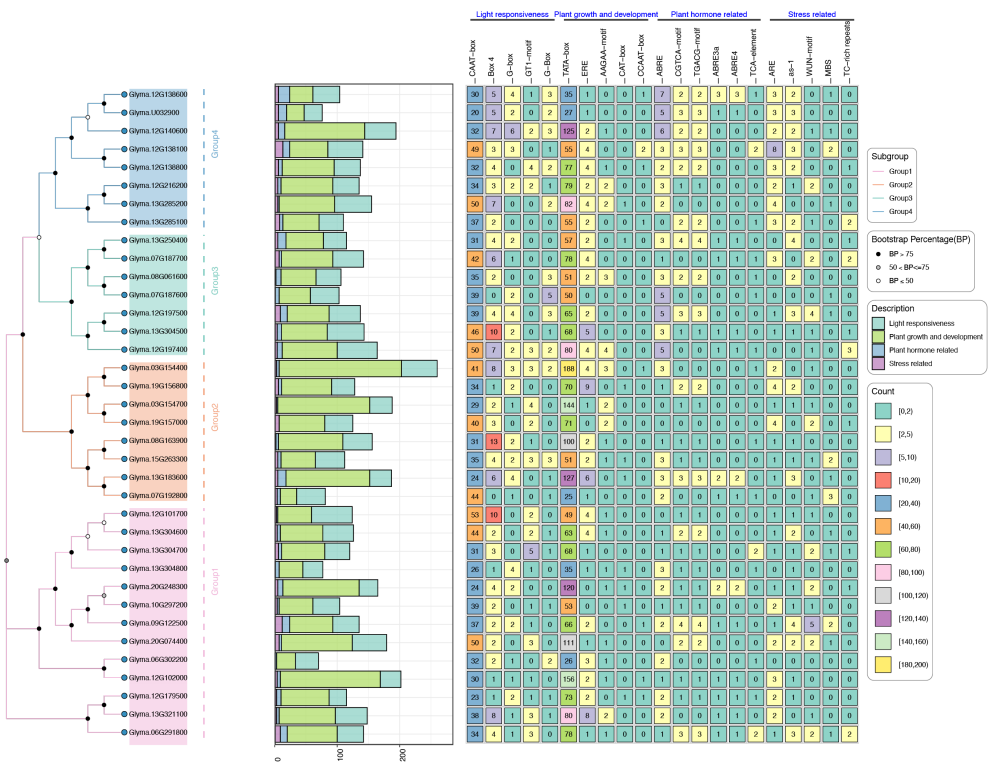


FIGURE 5
Cis-elements analysis of *TPS* genes in *G. max*.

10.3389/fpls.2024.1487092

AraNet2 database to analyze and predict a protein-protein interaction network. This analysis revealed extensive interconnectivity among the GmTPS proteins, with nearly all members engaging in interactions (Figure 6). Interestingly, we found that some GmTPS proteins can interact with P450 proteins, suggesting potential functional crosstalk or coordinated activities between these two protein classes (Figure 6).

In addition, we selected Glyma.07G192800 and Glyma.15G263300 proteins to verify whether TPS proteins can interact with other TPS proteins or P450 proteins. To this end, we used the Y2H assay. As data shown in Figure 7, we found that Glyma.07G192800 can interact with Glyma.12G140600, and Glyma.15G263300 can interact with Glyma.20G074400 (Figure 7). Moreover, we found that Glyma.07G192800 can interact with P450 proteins Glyma.09G029400, and Glyma.15G263300 can interact with P450 proteins Glyma.01G153300 (Figure 7). These data indicate that TPS protein may form heterodimer to function, or may form complex with P450 protein to function.

3.8 The tissues expression analysis of GmTPS

To further explore the functions of *GmTPS*, we analyzed its expression pattern based on RNA-seq data. *Glyma.06g45780*, *Glyma.12g16940*, *Glyma.12g32370* and *Glyma.07g30700*, expressed higher in flower than other tissues. We speculate that these GmTPS genes may synergistically regulate soybean growth and development (Figure 8). And Glyma.12g16940 and *Glyma.12g32370* are expressed only in flower. Additionally, these GmTPSs no expression of *Glyma.13g32380*, *Glyma.17g05500*, *Glyma.12g16990*, *Glyma.13g38050*, *Glyma.06g45780*, *Glyma.12g16940*, *Glyma.12g32370*, *Glyma.07g30700*, *Glyma.07g30710* and *Glyma.20g18280* are detected in the organs of seed and nodule. On the other hand, *Glyma.08g17470* and *Glyma.15g41670* are widely expressed in root, young leaf, pod shell, flower, seed and nodule (Figure 8). The result indicate that *GmTPS* genes may involve in diverse aspects of plant growth and developmental processes.

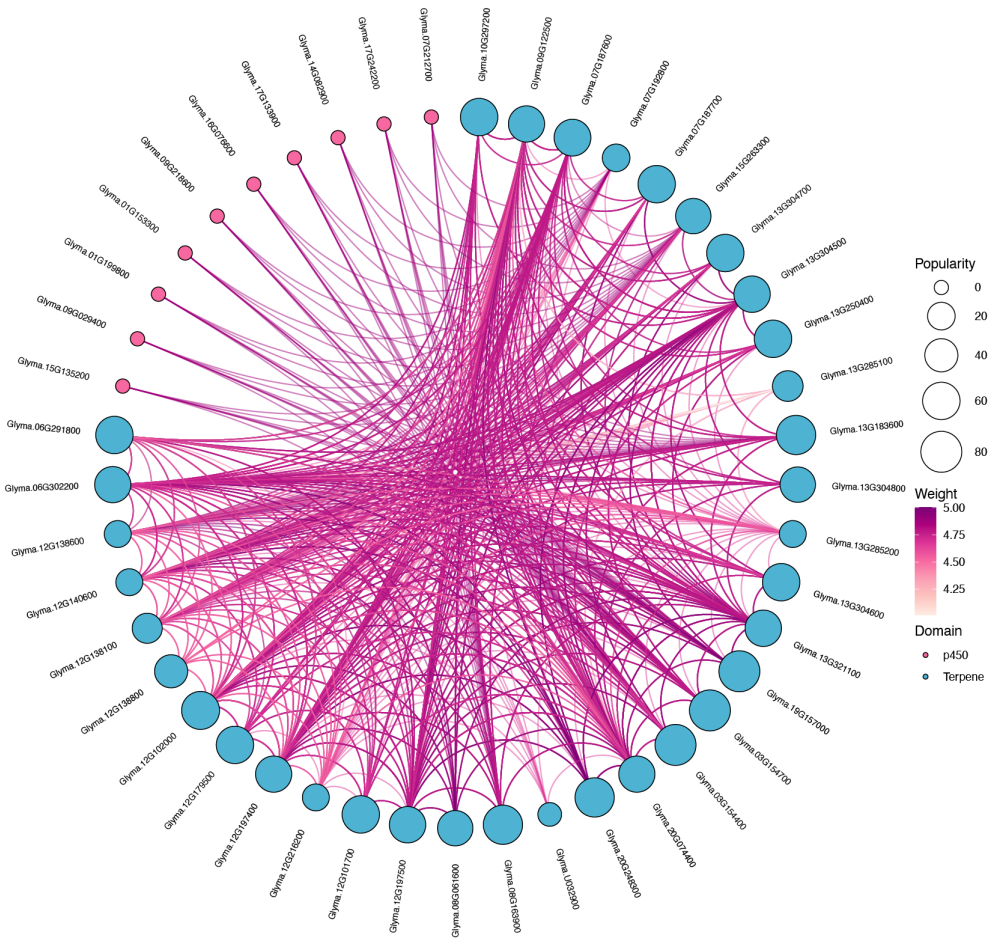


FIGURE 6
The protein-protein interaction network of TPSs protein in *G. max*.

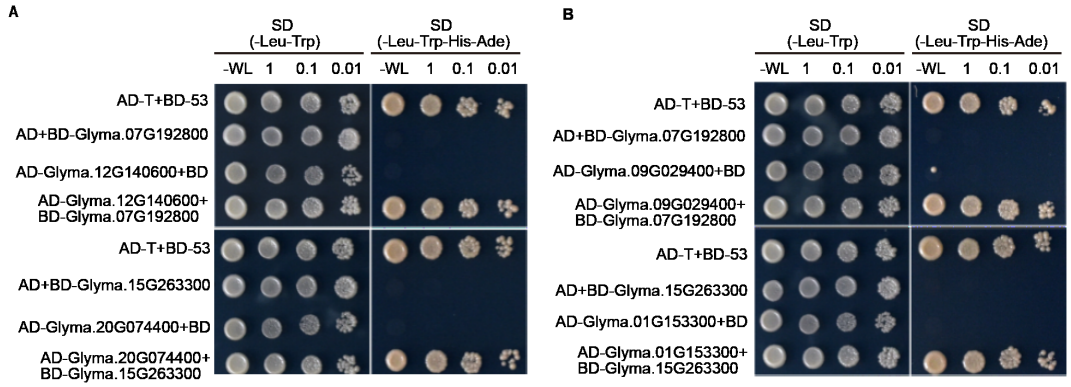


FIGURE 7
Y2H assays show TPS proteins can interact with other TPS proteins or P450 proteins. **(A)** The image showed the results of TPS proteins can interact with other TPS proteins. **(B)** The image showed the results of TPS proteins can interact with P450. The positive control: AD-T + BD-53.

4 Discussion

Terpenoids, widely existing in plants, fungi, bacteria and insects, and play an pivotal role in enhancing plant resistance. It is worth noting that the TPS proteins are involved in many biological processes, such as low-temperature stress adaptation, drought stress adaptation, salt stress adaptation, and responses to phytohormonal and insect resistance (Yeo et al., 2000; Li et al., 2011; Huang et al., 2018; Ge et al., 2008; Zhou et al., 2020; Yan et al., 2022; Zhong et al., 2024). While the *TPS* gene family has been found in

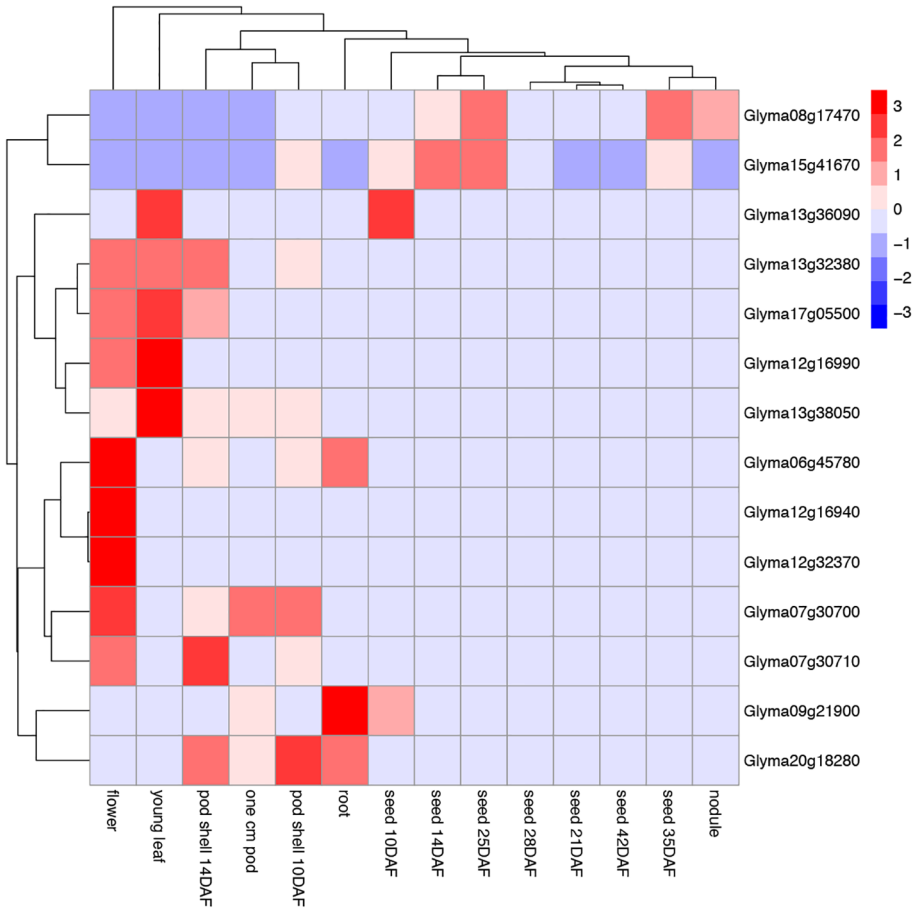


FIGURE 8
The heat map of the tissue expression of *GmTPS* genes in different tissues and development stages (flower, young leaf, pod shell 14DAF, one cm pod, pod shell 10DAF, root, seed 10DAF, seed 14DAF, seed 25DAF, seed 28DAF, seed 21DAF, seed 42DAF, seed 35DAF and nodule). The data are shown in a heatmap with gene expression in different tissues and development stages with row-scaled FPKM values.

many species, but, the whole-genome identification and bioinformatics analysis of *TPS* gene family in soybean is lacking. In this study, we systematically analyzed the *TPS* gene family in soybean using bioinformatics methods and identified a total of 36 *TPS* genes (Table 1). *TPS* genes, which are ubiquitous among plant species, such as 33, 53, 12, 8, 80, 26, 58 and 16 *TPS* genes were found in *Arabidopsis*, rice, *populus*, *potato*, *d. officinale*, *camellia*, *aloes*, *l. chinense* and *A. hypogaea* (Aubourg et al., 2002; Yang et al., 2012; Xu et al., 2017; Zhou et al., 2020; Li et al., 2021; Cao et al., 2023 and Zhong et al., 2024). It is not difficult to see that the member of TPS genes in different plants varies greatly. These results also showed that TPS genes may not only be functionally conserved, but also functionally differentiated in different plants.

Prior research has categorized TPS proteins into seven distinct subfamilies: TPS-a through TPS-h. are predominantly found in angiosperms, while TPS-c is specific to gymnosperms, and TPS-e/f occurs in vascular plants (Newman and Chappell, 1999; Chen et al., 2011). Phylogenetic tree showed that the 122 TPS proteins in GmTPSs, AtTPSs and OsTPSs can be divided into five groups (Figure 1). Interestingly, Group5 only exists in the rice, and Group1 and Group4 only exists in soybean and *Arabidopsis*. These observations may provide insights into the evolution and diversification of the *TPS* gene family in monocotyledons and dicotyledons.

During the progress of evolution, TD and WGD events played a key role in the expansion of gene families, new genes and novel functions (Freeling, 2009; Panchy et al., 2016). In our study, we observed rapid expansion of the Group4 subgroup in *G. max* due to recent TD and WGD, while the Group5 subgroup experienced rapid expansion in *O. sativa* for the TD. Overall, the number of WGD genes was the largest, indicating that WGD was found to be the predominant mechanism driving the evolution and expansion of the *TPS* gene family in *G. max*. This investigation offers profound insights into the evolutionary journey and expansion patterns of the *TPS* gene family across diverse plant species.

Numerous studies found the important role of *TPS* genes in mediating plant responses to different hormone signals, and abiotic and biotic stress (Yeo et al., 2000; Li et al., 2011; Huang et al., 2018; Ge et al., 2008; Zhou et al., 2020; Yan et al., 2022; Zhong et al., 2024). For instance, *OsTPS1* overexpression in rice boosts trehalose levels, enhancing resilience against low temperatures (Ge et al., 2008). Similarly, *TaTPS11* overexpression in *Arabidopsis* enhances cold tolerance (Liu et al., 2019), while *ScTPS1* overexpression in tomato elevates drought tolerance (Cortina and Culiáñez-Macià, 2005). Our cis-acting element analysis revealed that *Glyma.07G192800*, *Glyma.12G138100*, *Glyma.15G263300* and *Glyma.09G122500* contain three, two, two and two MBS cis-elements, respectively (Figure 5). This result implies that these *GmTPS* genes may be involved in drought signaling pathways, and will be an interesting topic to explore in the future.

On the other hand, MeJA treatment transcriptionally upregulated the expression of most *CsTPS* genes (Zhou et al., 2020). Our analysis uncovered the prevalence of MYC2, ABRE, CGTCA-motif, TGACG-motif, and WUN-motif in the promoters of *GmTPS* genes, particularly *Glyma.12G138100*, *Glyma.15G263300*, and *Glyma.09G122500*, which harbor ABRE, CGTCA-motif, TGACG-motif, as-1, WUN-motif, and MBS cis-elements (Figure 5). Despite the scarcity of experimental evidence elucidating the intricate relationships between phytohormone signaling and terpenes biosynthesis, we hypothesize that intricate crosstalks among distinct phytohormone signaling pathways may delicately modulate terpenes biosynthesis through a myriad of transcription factors, and will an interesting topic to explore in the future.

Cytochrome P450s (CYPs) orchestrate an array of essential processes, encompassing growth, development, and the biosynthesis of secondary metabolites (Mizutani and Ohta, 2010 and 2012). For instance, P450 enzymes exhibit remarkable adaptability in modulating plant development through hormone synthesis (Schuler, 1996). Specifically, *CYP707A* play a key role in the catalytic synthesis of ABA (Saito et al., 2004), while *CYP94B3*, *CYP94C1* and *CYP74B* were related in JA biosynthesis (Li et al., 2008; Koo et al., 2011; Heitz et al., 2012). Our observations that some TPS proteins can interact with P450 proteins in soybean during the Y2H assays (Figure 7), suggest that TPS proteins likely form complexes with P450 proteins and participate in the growth, development, and the biosynthesis of secondary metabolites in soybean.

5 Conclusions

In this study, we identified 36 *TPS* members in soybean and systematically grouped them into five distinct subfamilies: Group1, Group2, Group3, Group4 and Group5. Subsequently, we demonstrated that both TD and WGD contributed significantly to the expansion of TPS genes in *Glycine max*, with WGD playing a pivotal role. Furthermore, our analysis revealed that all *GmTPS*, *AtTPS*, and *OsTPS* genes were subjected to purifying selection. Yeast two-hybrid (Y2H) assay results showed that TPS protein may form a heterodimer to function, or may form a complex with P450 protein to function. RNA-seq data displayed *GmTPS* genes are involved in soybean growth and development. This exhaustive study establishes a foundational understanding of the pivotal roles played by *GmTPS* genes in soybean.

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

HL: Data curation, Formal analysis, Funding acquisition, Methodology, Writing – original draft, Writing – review & editing. XZ: Funding acquisition, Investigation, Methodology, Software, Writing – review & editing. QY: Methodology, Validation, Writing – review & editing. XS: Data curation, Investigation, Writing – review & editing. YM: Data curation, Formal analysis, Funding acquisition, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This project was supported by Shanxi Province Applied Basic Research Project (202103021224145, 202303021212083), Yuncheng Science and Technology Bureau Basic Research Project (YCKJ-2022073), Shanxi Agricultural University Cotton Research Institute Innovation Development project (SJJCX2023-04), Shanxi Province doctoral graduates, PhD After the researchers to Jin work award funding research project (SXBYKY2023024) support.

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.2024.1487092/full#supplementary-material

SUPPLEMENTARY DATA SHEET 1
The information of Ka, Ks and Ka/Ks values of duplicate gene pairs.

SUPPLEMENTARY DATA SHEET 2
The information on primer sequences of genes involved in this study.

References

Aubourg, S., Lecharny, A., and Bohlmann, J. (2002). Genomic analysis of the terpenoid synthase (*AtTPS*) gene family of *Arabidopsis thaliana*. *Mol. Genet. Genomics* 267, 730–745. doi: 10.1007/s00438-002-0709-y

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinf.* 10, 1–9. doi: 10.1186/1471-2105-10-421

Cao, Z. J., Ma, Q. X., Weng, Y. H., Shi, J. S., Chen, J. H., and Hao, Z. D. (2023). Genome-wide identification and expression analysis of *TPS* gene family in *Liriodendron chinense*. *Genes (Basel)* 14, 770. doi: 10.3390/genes14030770

Chen, F., Tholl, D., Bohlmann, J., and Pichersky, E. (2011). The family of terpene synthases in plants: a mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. *Plant J.* 66, 212–229. doi: 10.1111/j.1365-313X.2011.04520.x

Cortina, C., and Culiáñez-Macià, F. A. (2005). Tomato abiotic stress enhanced tolerance by trehalose biosynthesis. *Plant Sci.* 169, 75–82. doi: 10.1016/j.plantsci.2005.02.026

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acid Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340

Freeling, M. (2009). Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition. *Annu. Rev. Plant Biol.* 60, 433–453. doi: 10.1146/annurev.arplant.043008.092122

Ge, L. F., Chao, D. Y., Shi, M., Zhu, M. Z., Gao, J. P., and Lin, H. X. (2008). Overexpression of the trehalose-6-phosphate phosphatase gene *OsTPP1*confers stress tolerance in rice and results in the activation of stress responsive genes. *Planta* 228, 191–201. doi: 10.1007/s00425-008-0729-x

Heitz, T., Widemann, E., Lugan, R., Miesch, L., Ullmann, P., Désaubry, L., et al. (2012). Cytochromes P450 CYP94C1 and CYP94B3 catalyze two successive oxidation steps of plant hormone jasmonoyl-isoleucine for catabolic turnover. *J. Biol. Chem.* 287, 6296–6306. doi: 10.1074/jbc.M111.316364

Huang, X. Z., Xiao, Y. T., Köllner, T. G., Jing, W. H., Kou, J. F., Chen, J. W., et al. (2018). The terpene synthase gene family in Gossypium hirsutum harbors a linalool synthase *GhTPS12* implicated in direct defence responses against herbivores. *Plant Cell Environ.* 41, 261–274. doi: 10.1111/pce.13088

Jia, Q., Chen, X., Kollner, T. G., Rinkel, J., Fu, J., Labbe, J., et al. (2019). Terpene synthase genes originated from bacteria through horizontal gene transfer contribute to terpenoid diversity in Fungi. *Sci. Rep.* 9, 9223. doi: 10.1038/s41598-019-45532-1

Kaul, S., Koo, H. L., Jenkins, J., Rizzo, M., and Rooney, T. (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796–815. doi: 10.1038/35048692

Kolde, R., and Kolde, M. R. (2015). PHeatmap: Pretty Heatmaps. *R Package.* 1, 1. Available online at: https://cran.r-project.org/web/packages/pheatmap/pheatmap.pdf.

Koo, A. J. K., Cooke, T. F., and Howe, G. A. (2011). Cytochrome P450 CYP94B3 mediates catabolism and inactivation of the plant hormone jasmonoyl-L-isoleucine. *Proc. Natl. Acad. Sci.* 108, 9298–9303. doi: 10.1073/pnas.1103542108

Lee, T., Yang, S., Kim, E., Ko, Y., Hwang, S., Shin, J., et al. (2015). AraNet v2: an improved database of co-functional gene networks for the study of *Arabidopsis thaliana* and 27 other nonmodel plant species. *Nucleic Acid Res.* 43, D996–D1002. doi: 10.1093/nar/gku1053

Lescot, M., Déhais, P., Thijs, G., Marchal, K., Moreau, Y., and Van, de Peer, Y. (2002). PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acid Res.* 30, 325–327. doi: 10.1093/nar/30.1.325

Li, L., Chang, Z., Pan, Z., Fu, Z.-Q., and Wang, X. (2008). Modes of heme binding and substrate access for cytochrome P450 CYP74A revealed by crystal structures of allene oxide synthase. *Proc. Natl. Acad. Sci.* 105, 13883–13888. doi: 10.1073/pnas.0804099105

Li, H. W., Zang, B. S., Deng, X. W., and Wang, X. P. (2011). Overexpression of the trehalose-6-phosphate synthase gene *OsTPS1* enhances abiotic stress tolerance in rice. *Planta* 234, 1007–1018. doi: 10.1007/s00425-011-1458-0

Li, R. S., Zhu, J. H., Guo, D., Li, H. L., Wang, Y., Ding, X. P., et al. (2021). Genome-wide identification and expression analysis of terpene synthase gene family in Aquilaria sinensis. *Plant Physiol. Biochem.* 164, 185–194. doi: 10.1016/j.plaphy.2021.04.028

Lin, J. Y., Wang, D., Chen, X. L., Kollner, T. G., Mazarei, M., Guo, H., et al. (2017). An (E,E)-alpha-farnesene synthase gene of soybean has a role in defence against nematodes and is involved in synthesizing insect-induced volatiles. *Plant Biotechnol. J.* 15, 510–519. doi: 10.1111/pbi.12649

Liu, X., Fu, L., Qin, P., Sun, Y., Liu, J., and Wang, X. (2019). Overexpression of the wheat trehalose 6-phosphate synthase 11 gene enhances cold tolerance in *Arabidopsis thaliana*. *Gene* 710, 210–217. doi: 10.1016/j.gene.2019.06.006

Mistry, J., Finn, R. D., Eddy, S. R., Bateman, A., and Punta, M. (2013). Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acid Res.* 41, e121–e121. doi: 10.1093/nar/gkt263

Mizutani, M. (2012). Impacts of diversification of cytochrome P450 on plant metabolism. *Biol. Pharm. Bull.* 35, 824–832. doi: 10.1248/bpb.35.824

Mizutani, M., and Ohta, D. (2010). Diversification of P450 genes during land plant evolution. *Annu. Rev. Plant Biol.* 61, 291–315. doi: 10.1146/annurev-arplant-042809-112305

Newman, J. D., and Chappell, J. (1999). Isoprenoid biosynthesis in plants: carbon partitioning within the cytoplasmic pathway. *Crit. Rev. Biochem. Mol. Biol.* 34, 95–106. doi: 10.1080/10409239991209228

Nguyen, L.-T., Schmidt, H. A., Von, Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300

Panchy, N., Lehti,-Shiu, M., and Shiu, S. H. (2016). Evolution of gene duplication in plants. *Plant Physiol.* 171, 2294–2316. doi: 10.1104/pp.16.00523

Paysan, L. T., Blum, M., Chuguransky, S., Grego, T., Pinto, B. L., Salazar, G. A., et al. (2023). InterPro in 2022. *Nucleic Acids Res.* 51, D418–D427. doi: 10.1093/nar/gkac993

Realdon, E. (1960). Modern classification of the terpenoids. *Boll. Chim. Farm* 99, 52–58.

Saito, S., Hirai, N., Matsumoto, C., Ohigashi, H., Ohta, D., Sakata, K., et al. (2004). Arabidopsis CYP707As encode (+)-abscisic acid 8′-hydroxylase, a key enzyme in the oxidative catabolism of abscisic Acid. *Plant Physiol.* 134, 1439–1449. doi: 10.1104/pp.103.037614

Schuler, M. A. (1996). Plant cytochrome P450 monooxygenases. *Crit. Rev. Plant Sci.* 15, 235–284. doi: 10.1080/07352689609701942

Sun, Y., Huang, X., Ning, Y., Jing, W., Bruce, T. J., Qi, F., et al. (2017). TPS46, a rice terpene synthase conferring natural resistance to bird cherry-Oat aphid, rhopalosiphum padi *(Linnaeus*). *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.00110

Tang, H., Bowers, J. E., Wang, X., Ming, R., Alam, M., and Paterson, A. H. (2008). Synteny and collinearity in plant genomes. *Science* 320, 486–488. doi: 10.1126/science.1153917

Thompson, J. D., Gibson, T. J., and Higgins, D. G. (2003). Multiple sequence alignment using ClustalW and ClustalX. *Curr. Protoc. Bioinf.* 2, 312–322. doi: 10.1002/0471250953.bi0203s00

Wang, S., Ouyang, K., and Wang, K. (2019). Genome-wide identification, evolution, and expression analysis of TPS and *TPP* gene families in *Brachypodium distachyon*. *Plants* 8, 362. doi: 10.3390/plants8100362

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

Xu, Y., Wang, Y., Mattson, N., Yang, L., and Jin, Q. (2017). Genome-wide analysis of the *Solanum tuberosum* (potato) trehalose-6-phosphate synthase (*TPS*) gene family: evolution and differential expression during development and stress. *BMC Genomics* 18, 1–15. doi: 10.1186/s12864-017-4298-x

Yan, Y., Li, M., Zhang, X., Kong, W., Bendahmane, M., Bao, M., et al. (2022). Tissue-specific expression of the terpene synthase family genes in *Rosa chinensis* and effect of abiotic stress conditions. *Genes* 13, 547. doi: 10.3390/genes13030547

Yang, H. L., Liu, Y. J., Wang, C. L., and Zeng, Q. Y. (2012). Molecular evolution of trehalose-6-phosphate synthase (*TPS*) gene family in Populus, Arabidopsis and rice. *PloS One* 7, e42438. doi: 10.1371/journal.pone.0042438

Yeo, E. T., Kwon, H. B., Han, S. E., Lee, J. T., Ryu, J. C., and Byu, M. (2000). Genetic engineering of drought resistant potato plants by introduction of the trehalose-6-phosphate synthase (*TPS1*) gene from *Saccharomyces cerevisiae*. *Molecules Cells* 10, 263–268. doi: 10.1016/S1016-8478(23)17473-5

Yoshitomi, K., Taniguchi, S., Tanaka, K., Uji, Y., Akimitsu, K., and Gomi, K. (2016). Rice terpene synthase 24 (*OsTPS24*) encodes a jasmonate-responsive monoterpene synthase that produces an antibacterial gamma-terpinene against rice pathogen. *J. Plant Physiol.* 191, 120–126. doi: 10.1016/j.jplph.2015.12.008

Yu, G., Smith, D. K., Zhu, H., Guan, Y., and Lam, T. T. Y. (2017). Ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* 8, 28–36. doi: 10.1111/mee3.2017.8.issue-1

Yu, Z., Zhao, C., Zhang, G., Teixeira, da Silva, J. A., and Duan, J. (2020). Genome-wide identification and expression profile of *TPS* gene family in *Dendrobium officinale* and the role of *DoTPS10* in linalool biosynthesis. *Int. J. Mol. Sci.* 21, 5419. doi: 10.3390/ijms21155419

Zhang, Z. (2022). KaKs_Calculator 3.0: calculating selective pressure on coding and non-coding sequences. *Genomics Proteomics Bioinf.* 20, 536–540. doi: 10.1016/j.gpb.2021.12.002

Zhong, C., He, Z., Liu, Y., Li, Z., Wang, X., Jiang, C., et al. (2024).). Genome-wide identification of *TPS* and *TPP* genes in cultivated peanut (*Arachis hypogaea*) and functional characterization of *AhTPS9* in response to cold stress. *Front. Plant Sci.* 14. doi: 10.3389/fpls.2023.1343402

Zhou, H. C., Shamala, L. F., Yi, X. K., Yan, Z., and Wei, S. (2020). Analysis of terpene synthase family genes in Camellia sinensis with an emphasis on abiotic stress conditions. *Sci. Rep.* 10, 933. doi: 10.1038/s41598-020-57805-1

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 27 September 2024
DOI 10.3389/fpls.2024.1479536

OPEN ACCESS

EDITED BY
Zhenbin Hu,
Agricultural Research Service (USDA), United States

REVIEWED BY
Yuzhou Xu,
Kansas State University, United States
Zixiang Wen,
Syngenta, United States

*CORRESPONDENCE
Huatao Chen
cht@jaas.ac.cn
Xuejun Wang
wangxj4002@sina.com

†These authors have contributed equally to this work

RECEIVED 12 August 2024
ACCEPTED 09 September 2024
PUBLISHED 27 September 2024

CITATION
Jia Q, Hu S, Li X, Wei L, Wang Q, Zhang W, Zhang H, Liu X, Chen X, Wang X and Chen H (2024) Identification of candidate genes and development of KASP markers for soybean shade-tolerance using GWAS.
Front. Plant Sci. 15:1479536.
doi: 10.3389/fpls.2024.1479536

Identification of candidate genes and development of KASP markers for soybean shade-tolerance using GWAS

Qianru Jia[1†], Shengyan Hu[1†], Xihuan Li[1], Libin Wei[2], Qiong Wang[1], Wei Zhang[1], Hongmei Zhang[1], Xiaoqing Liu[1], Xin Chen[1], Xuejun Wang[2*] and Huatao Chen[1,3*]

[1]Institute of Industrial Crops, Jiangsu Academy of Agricultural Sciences, Nanjing, China, [2]Jiangsu Yanjiang Institute of Agricultural Sciences, Nantong, China, [3]Zhongshan Biological Breeding Laboratory (ZSBBL), Nanjing, China

Shade has a direct impact on photosynthesis and production of plants. Exposure to shade significantly reduces crops yields. Identifying shade-tolerant genomic loci and soybean varieties is crucial for improving soybean yields. In this study, we applied a shade treatment (30% light reduction) to a natural soybean population consisting of 264 accessions, and measured several traits, including the first pod height, plant height, pod number per plant, grain weight per plant, branch number, and main stem node number. Additionally, we performed GWAS on these six traits with and without shade treatment, as well as on the shade tolerance coefficients (STCs) of the six traits. As a result, we identified five shade-tolerance varieties, 733 SNPs and four candidate genes over two years. Furthermore, we developed four kompetitive allele-specific PCR (KASP) makers for the STC of S18_1766721, S09_48870909, S19_49517336, S18_3429732. This study provides valuable genetic resources for breeding soybean shade tolerance and offers new insights into the theoretical research on soybean shade tolerance.

KEYWORDS

soybean, shade tolerance, shade tolerance coefficient, GWAS, KASP

Introduction

Weak or low light conditions reduce the capacity of photosynthesis, which can ultimately lead to plant starvation and cause a series of disruptions in the physiological and biochemical metabolic processes throughout the plant's entire life cycle. These disruptions include leaf curling and thinning, loss of greenery, premature leaf senescence, reduced branching, slow growth, decreased resistance, and lower plant yield and biomass (Li et al., 2012; Fankhauser and Batschauer, 2016; Li et al., 2023b; Martinez-Garcia and Rodriguez-Concepcion, 2023).

Soybean is a photophilic crop with a high demand for sunlight. However, to increase cultivation area and yield, soybeans are often interplanted with maize, sorghum, sunflower or fruit trees (Echarte et al., 2011; Ghosh et al., 2009; Yang et al., 2015; Su et al., 2023a). Under intercropping or high-density planting conditions, soybeans experience shade stress, which negatively impacts their yield and quality (Gong et al., 2015; Wu et al., 2017; Su et al., 2023a). Research has shown that weak light conditions can reduce the photosynthetic rate and chlorophyll a/chlorophyll b ration in soybean leaves, leading to a decline in photosynthetic capacity (Su et al., 2023a). Using gene/allele sequence markers (GASM-RTM-GWAS), Su et al. identified 140 genes or alleles associated with the shade-tolerance index (STI), 146 with relative pith cell length (RCL), and nine with both (Su et al., 2023b). Through transcriptome and metabolome sequence analysis of the shade-tolerant soybean 'Nanxiadou 25' under natural and 50% light conditions, 36 differentially expressed genes and 12 potential candidate genes related to shade tolerance were identified, including ATP phosphoribosyl transferase, phosphocholine phosphatase, AUXIN-RESPONSIVE PROTEIN, PURPLE ACID PHOSPHATASE (Jiang et al., 2023). Nandou 12 has demonstrated stronger shade resistance and a quicker recovery compared to Jiuyuehuang (shade-intolerant) during light recovery, due to its higher photosynthetic rate and smaller decrease in soluble sugar and protein content (Wang et al., 2023). Li et al. identified 29 up-regulated and 412 down-regulated proteins in soybeans seedlings exposed to 2-hour shade stress compared to those under white light. They also found that shade stress significantly impacted carbohydrate metabolic processes, especially cell wall polysaccharide biosynthetic pathways (Li et al., 2019b).

Key genes related to various agronomic traits that influence soybean shade tolerance or adaptability to high-density planting have also been identified. *PH13*, which encodes a WD40 protein and was identified through GWAS. The deletion of both the PH13 and its paralogue PHP can prevent shade-induced excessive stem elongation and enable high-density planting (Qin et al., 2023). The RIN1 (reduced internode 1) interacts with ELONGATED HYPOCOTYL 5 (HY5), STF1 and STF2 to regulate gibberellin metabolism, which controls internode length. Mutations of *RIN1* result in shorter internodes and can enhance yield in high-density planting conditions (Li et al., 2023a).

Notably, previous studies have predominantly focused on shade tolerance, which falls short of addressing the full spectrum production needs. In our study, we treated a natural population of 264 soybean accessions with a 30% reduction in light to assess their response to shade. We used the shade tolerance coefficient (STC) as the evaluation metric. Through genome-wide association study (GWAS), we identified SNPs and candidate genes associated with shade tolerance. We developed KASP markers for S18_1766721, S09_48870909, S19_49517336, S18_3429732, which have been successfully applied. This research provides new insights into the development of shade-tolerant soybean germplasm and offers valuable resources for cultivation strategies.

Materials and methods

Materials

A natural population consisting of 264 Chinese soybean accessions, including 212 improved varieties and 52 landraces, was utilized in this study. Genome-wide association study of the landrace panel and the cultivated panel was conducted with 2,597,425 SNPs. The particular information has been presented in our previous research (Zhang et al., 2021).

Shade treatment and shade-tolerance evaluation

The shading stress was simulated using shade nets that reduced light by 30%, and the results were compared to normal conditions with natural light. The study took place in Nantong (32°1'N, 120° 52'E), Jiangsu Province, China. Soybean germplasms were planted in June and harvested in October of both 2022 and 2023. Each soybean germplasm material was grown in 3 rows, with 10 holes per row, and each row is filled with 20-25 plants. After harvesting, six traits (first pod height, plant height, pod number per plant, grain weight per plant, branch number and main stem node number) were measured based on the *Descriptors and Date Standard for Soybean (Glycine* spp.*)* (Qiu and Chang, 2006). The shade tolerance coefficient (STC) for each trait was used as an evaluation indicator, calculated using the following formula:

$$\mathrm{STC}_{ij} = \bar{y}_{ij(Treat)} / \bar{y}_{ij(CK)} \times 100\,\%$$

In which, $\bar{y}_{ij(Treat)}$ and $\bar{y}_{ij(CK)}$ represent average observed value of genotype i (i=1, 2, 3…264) on the trait j (j=1, 2, 3) with or without shade treatment (Li et al., 2014).

Standardize the STC of each genotype for each trait using the subordinate function value (SFV) (scaled to the interval ([0,1]) by the following formulas:

$$F_{ij} = \frac{(STC_{ij} - \min(STCij))}{(\max(STC_{ij}) - \min(STC_{ij}))}$$

$$W_j = pj / \sum_{j=1}^{n} pj$$

$$D = \sum_{j=1}^{n} [u(X_j) * W_j]$$

In which, min (STCij) and max (STCij) represent the j (j=1, 2, 3) trait minimum and maximum of genotype i (i=1, 2, 3… 264), respectively; W_j represents the importance or weight of the j trait among all composite indicators, where *pj* is the contribution rate of the j trait for each soybean genotype. The D value is the comprehensive evaluation score of shade tolerance for each soybean genotype under shade stress conditions, obtained by assessing the comprehensive indicators. Here, Xj represents the j trait (Li et al., 2014; Du, 2023).

Based on the calculated Average Subordinate Function Value (ASFV), the data for all genotypes under the current trait are evenly divided into five categories. The grouping criteria for shade tolerance in these five categories are determined, with each genotype being classified according to its shade tolerance level. The higher the ASFV, the stronger the shade tolerance of the genotype.

GWAS

The population resequencing data utilized in this study was previously reported in our earlier research. In brief, high-density map includes 2,597,425 single nucleotide polymorphisms (SNPs) from the landrace and cultivated accessions, with a linkage disequilibrium (LD) decay range of 120 kb (Zhang et al., 2021). For each year of the study, ten plants were selected for measurement. Genome-Wide Association Studies (GWAS) were conducted using the GAPIT package based on R software and a mix linear model (MLM) were employed.

KASP

Genotyping was performed using three sets of primers (F1, F2, and R) specifically designed for KASP markers, as detailed in Supplementary Table S3. These primers were designed using the Primer-Blast tool available on the NCBI website (https://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi?LINK_LOC=BlastHome). Genomic DNA was extracted using the 2×CTAB method (Jia et al., 2024). PCR amplification was carried out using the KASP V4.0 2×Mastermix (JasonGen, China), following the reagent's instructions. The amplified DNA was then analyzed using a Quantitative Real-Time PCR System (ABI Quant Studio 5).

Quantification and statistical analysis

The software IBM SPSS 20 was used for descriptive statistics and analysis of variance (ANOVA) (IBM, Armonk, NY, USA). Correlation analyses were performed using Origin software (Origin Lab, USA). The frequency distributions of six traits for the soybean accessions in both years were calculated by Microsoft Excel 2016.

Results

Shade-tolerance evaluation and analysis of the six agronomic traits across the 264 soybean accessions with or without shade treatment

To evaluate shade tolerance, we cultivated a natural soybean population of 264 accessions under conditions with or without 30% shade treatment in Nantong during 2022 and 2023. Six agronomic traits (first pod height, plant height, main stem node number, pod number per plant, grain weight per plant, and branch number) were measured across the 264 accessions over two years. Overall, the average first pod height and plant height in 2022 (E1) and 2023 (E2) under shade treatment were higher than those under normal light conditions. In contrast, the average main stem node number, pod number per plant, grain weight per plant, and branch number showed varying degrees of decline (Supplementary Table S1). To evaluate the shade tolerance of the soybean population, STC for six traits was calculated, and descriptive statistical analysis were performed to the 264 accessions from 2022 and 2023 (Table 1). The STC values for six traits were defined as STC1-6, respectively. As shown, the mean STC values for these six traits in 2022 and 2023 did not exhibit significant differences (Table 1), with heritability (h^2)

TABLE 1 Descriptive statistics of STCs of six traits across 264 soybean accessions with or without shade treatment.

Trait	Year	Max	Min	Mean		SD	CV (%)	Skewness	Kurtosis	h^2 (%)
STC1	E1	7.00	0.36	1.74	1.76	1.04	59.43	1.817	5.085	43.64
	E2	5.19	0.59	1.77		0.77	43.65	1.278	2.68	
STC2	E1	8.54	0.76	1.92	1.91	0.80	41.38	3.215	20.50	40.15
	E2	3.61	0.23	1.90		0.63	32.96	0.417	-0.050	
STC3	E1	1.96	0.56	1.05	1.05	0.25	24.11	0.873	0.737	44.57
	E2	1.99	0.51	1.04		0.27	25.52	0.516	0.057	
STC4	E1	2.42	0.19	0.90	0.98	0.39	43.72	1.171	1.924	36.52
	E2	4.29	0.25	1.06		0.51	48.20	2.114	8.463	
STC5	E1	6.97	0.00	0.93	1.09	0.68	73.17	3.646	26.463	40.61
	E2	5.28	0.22	1.24		0.74	60.20	2.222	7.798	
STC6	E1	10.00	0.00	1.09	1.18	1.43	130.92	3.823	17.216	38.99
	E2	12.33	0.00	1.26		0.98	77.83	6.931	73.241	

STC1, STC of first pod height; STC2, STC of plant height; STC3, STC of node number on main stem; STC4, STC of pod number per plant; STC5, STC of seed weight; STC6, STC of branch number. Max, maximum; Min, minimum; SD, standard deviation; CV, coefficient of variation; h^2, heritability.

values of 43.64%, 40.15%, 44.57%, 36.52%, 40.61% and 38.99%, respectively (Table 1).

An analysis of variance (ANOVA) was conducted on the six traits across the 264 accessions in 2022 and 2023, revealing significant differences among genotypes, stress treatments, and different environments (Table 2). To explore the correlation among the six traits in 2022 and 2023 for the soybean population, a correlation analysis was conducted (Figure 1). The results indicated that the STC1 and STC2, STC2 and STC3, STC3 and STC4, STC4 and STC6 showed significant positive correlation in 2022 and 2023 (Figure 1). While, STC1 and STC4 in 2022, STC1 and STC4, STC5 in 2022, STC2 and STC5 in 2023 exhibited significant negative correlation (Figure 1). Between two years, only STC4 in 2022 and STC2 in 2023, STC1 in 2022 and STC5 in 2023 shown negative correlation, STC5 in 2022 and STC4 in 2023, STC4 in 2022 and STC5 in 2023 shown significant positive correlation, respectively. But other traits between two years shown weak significant correlation (Figure 1).

Shade tolerance soybean germplasms

In 2022, the ASFV thresholds for different levels of shade tolerance were as follows: high shade tolerance was above 0.68, shade tolerance ranged from 0.47 to 0.68, moderate shade tolerance ranged from 0.38 to 0.47, shade sensitivity ranged from 0.29 to 0.38, and high shade sensitivity was below 0.29 (Figure 2A). In 2023, the thresholds were slightly adjusted: high shade tolerance was above 0.70, shade tolerance ranged from 0.56 to 0.70, moderate shade tolerance ranged from 0.49 to 0.56, shade sensitivity ranged from 0.41 to 0.49, and high shade sensitivity was below 0.41 (Figure 2B). Over the two years, moderate shade tolerant soybean germplasm was the most prevalent, comprising approximately 39% and 41% of the total population. Shade sensitive germplasm ranked second after moderate shade tolerant germplasm. High shade tolerant germplasm was relatively rare, accounting for 0.76% and 1.52% of the total soybean population in 2022 and 2023, respectively (Figure 2A).

In summary, this study identified a total of five high shade tolerant soybean germplasms over the two years, with NPS044 being selected in both years (Table 3). These high shade tolerant materials offer a valuable foundation for further research into the genetic mechanisms underlying soybean shade tolerance and serve as important experimental materials for future breeding programs aimed at enhancing shade tolerance in soybeans.

GWAS for six agronomic traits and STCs across the 264 soybean accessions with or without shade treatment

To pinpoint key genomic loci responsible for shade tolerance in soybeans, we conducted GWAS on six traits across 264 soybean accessions under control and shade conditions, as well as STC of the

TABLE 2 Variance analysis of six traits in soybean natural population.

Trait	Variation source	Square Sum	Mean Square	F value	P value
First pod height	G	900.129	4.018	2.241	<0.01
	E	0.286	0.286	0.159	0.69
	G×E	760.974	3.397	1.895	<0.01
Plant height	G	421.892	1.883	2.582	<0.01
	E	0.509	0.509	0.698	0.404
	G×E	465.538	2.078	2.849	<0.01
Stem node number	G	69.076	0.308	3.05	<0.01
	E	0.214	0.214	2.121	0.146
	G×E	63.242	0.282	2.792	<0.01
Pod number per plant	G	161.9	0.723	1.725	<0.01
	E	0.041	0.041	0.097	0.756
	G×E	187.62	0.838	1.999	<0.01
Seed weight per plant	G	133.354	0.595	1.246	0.027
	E	7.309	7.309	15.294	<0.01
	G×E	87.727	0.392	0.82	0.954
Branch number	G	307.1	1.371	1.848	<0.01
	E	0.007	0.007	0.009	0.923
	G×E	314.363	1.403	1.892	<0.01

G, genotype; E, environment; SS, square sum; MS, mean square.

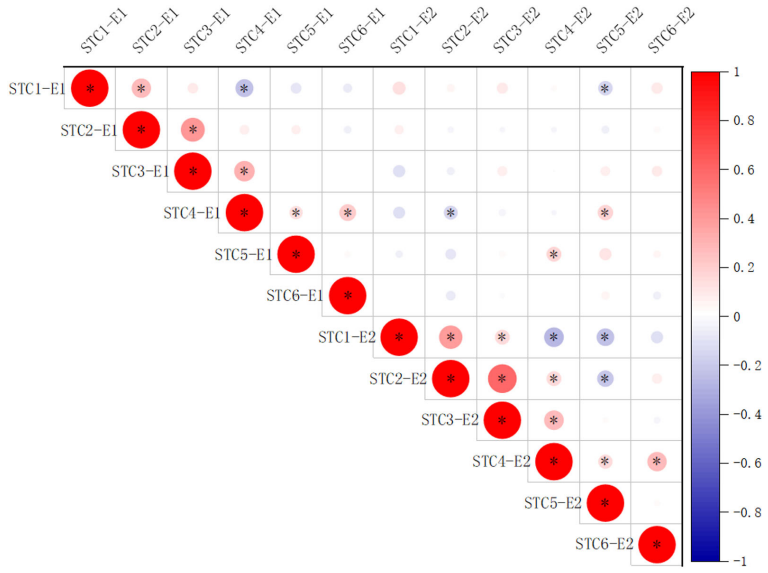


FIGURE 1
Correlation analysis among STC1, STC2, STC3, STC4, STC5, and STC6 of 2022 and 2023. E1, 2022; E2, 2023. *Represents $P<0.05$.

six traits for the years 2022 and 2023 (Figures 3–6; Supplementary Figures S3, S4). The resulting frequency distribution maps and density curves indicated that the phenotypic data for the six traits followed a continuous distribution. This suggests that the natural soybean population in our study harbors rich genetic variation, making it well-suited for further GWAS analyses (Supplementary Figures S1, S2).

Over the course of two years, we identified a total of 733 significant SNPs associated with STCs of six traits (Table 4). Specifically, for the STC of first pod height, 28 SNPs were detected. In 2022, 24 SNPs were associated with the STC of first pod height, while in 2023, 4 SNPs were identified. Notably, S11_19943066 was significant under control conditions, whereas S11_19738980 was significant under shade, with these two SNPs being approximately 204 kb apart (Figures 3A, B; Table 4). In 2023, S15_50935714 was significant under shade treatment, while S15_51517560 showed a significant correlation with the STC, with a distance of approximately 582 kb between these two SNPs (Figures 3E, F; Table 4).

In 2022, 29 SNPs were linked to the STC of plant height, while in 2023, 9 SNPs were identified. The SNP S15_15344441 showed significant association under control conditions, and S15_19750516 was significantly correlated with the STC (Figures 4A, B; Table 4). Moreover, S19_45102497 and S19_45149787 were significantly associated with plant height under control and shade conditions, respectively, with a distance of approximately 47 kb between them. The SNP S14_4288867 was significantly correlated with both plant height and STC. Additionally, S15_36102679 and S15_36134614 were significantly associated with soybean plant height under shading conditions and STC, with these two SNPs being approximately 32 kb apart (Figures 4D–F; Table 4).

For the main stem node number, two SNPs showed significant correlation with the STC. The SNPs S04_11807969 and

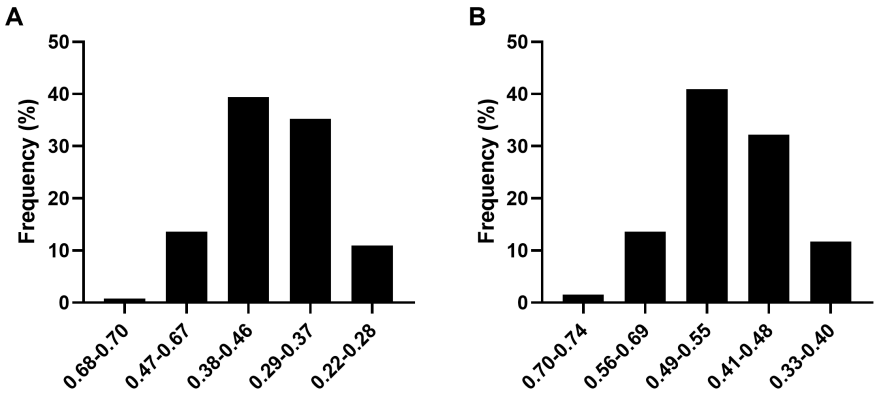


FIGURE 2
ASFV of 2022 **(A)** and 2023 **(B)**.

10.3389/fpls.2024.1479536

TABLE 3 Shade-tolerant soybean germplasms that were screened in 2022 and 2023.

Classification	High shade tolerance	ASFV
2022	NPS044	0.70
	NPS060	0.69
2023	NPS044	0.74
	NPS187	0.73
	NPS254	0.73
	NPS151	0.72

S14_4288867 were notably correlated with both the main stem node number under shading and the STC. Additionally, 2 SNPs were significantly correlated with the STC in 2023 (Supplementary Figure S3; Table 4).

For the trait of pod number per plant, 13 SNPs were identified as significantly correlated with the STC of pod number per plant in 2022. Additionally, 31 SNPs were significantly associated with the STC (Supplementary Figure S4; Table 4).

Regarding grain weight per plant, a total of 105 SNPs in 2022 and 190 SNPs in 2023 were significantly correlated with the STC. Specifically, the SNP S15_1302099 showed a significant correlation with both single plant grain weight and the STC of soybean under shading. The SNP S10_43122500 was significantly associated with grain weight per plant under shading conditions, while S10_43121286 was significantly correlated with the STC, with these two SNPs being approximately 1 kb apart. Notably, the SNP associated with the STC for grain weight per plant had the highest explanatory power, with a -$\log_{10}$(p) value of 9.28 and a phenotype explanatory rate of 18.02% (Figure 5; Table 4).

A total of 324 SNPs were found to be significantly correlated with the STC of branch number. Specifically, 227 SNPs in 2022 and 97 SNPs in 2023 were significantly associated with the STC (Table 4). In 2023, the SNP S18_55349193 showed significant correlation with branch number under control conditions, while S18_55354172 was significantly correlated with the STC, with these two SNPs being approximately 5 kb apart (Figure 6).

Development and application of KASP markers for soybean shade tolerance

To explore the phenotypic effects of allelic variations in significant SNPs, a haplotype analysis was conducted on the SNPs with the highest threshold detected for shade tolerance during the mature stages of 2022 and 2023. This analysis revealed a total of 4 SNPs showing significant differences between each genotype. For instance, the SNP S18_1766721 exhibited an allelic variation from A to G. The STC of first pod height was significantly higher in germplasm carrying the S18_1766721-G allele compared to those with the S18_1766721-A allele (Figure 7A). Another example includes the S09_48870909, which has a G/T allelic variation (Figure 7B). The nucleotide change at position S19_49517336 involves a substitution from G to A. Soybean with the S19_49517336-G allele exhibit a significantly higher average STC of pod number per plant compared to those with the S19_49517336-A allele (Figure 7C). For the S18_3429732, the allelic variation consists of A and G. Soybean germplasm with the S18-3429732-A allele has a significantly higher STC for average grain weight per plant compared to those with the S18-3429732-G allele

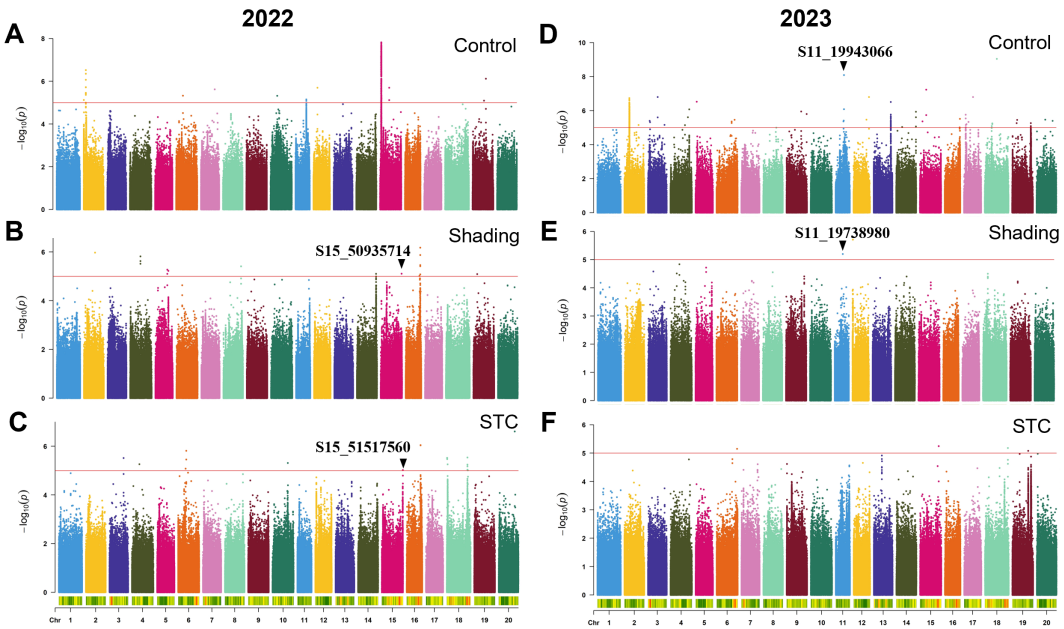


FIGURE 3
GWAS for first pod height with or without shade treatment in 2022 and 2023. **(A–C)** control, shade treatment and STC of 2022; **(D–F)** control, shade treatment and STC of 2023, respectively. Red lines represent-log10(p)≥5.0.

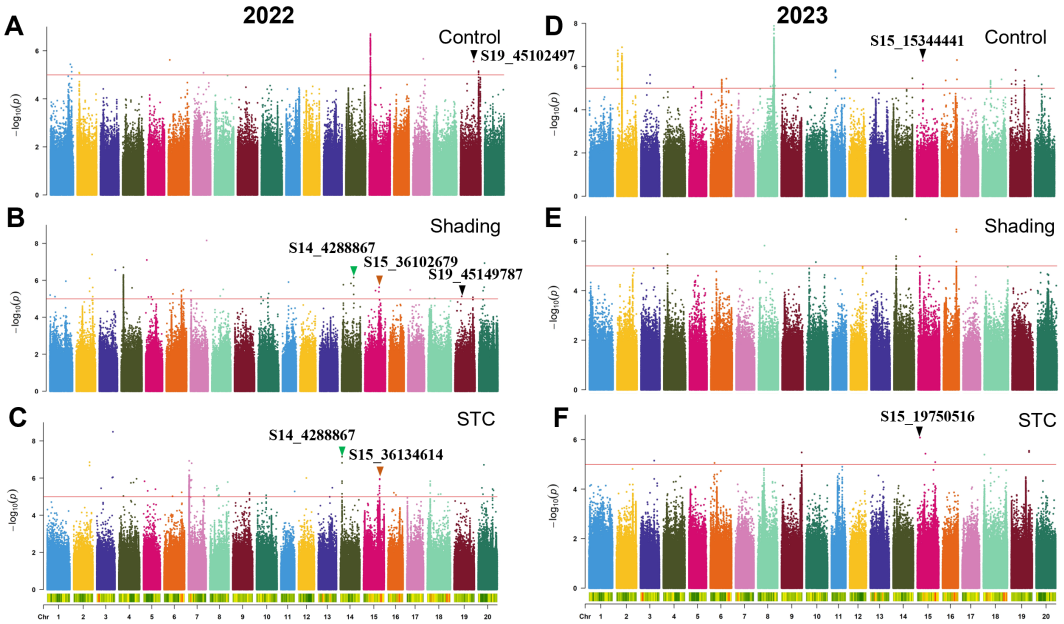


FIGURE 4
GWAS for plant height with or without shade treatment in 2022 and 2023. **(A–C)**, control, shade treatment and STC of 2022; **(D–F)**, control, shade treatment and STC of 2023, respectively. Red lines represent $-\log_{10}(p) \geq 5.0$.

(Figure 7D). The phenotypic variation explain rate of S18_1766721, S09_48870909, S19_49517336, and S18_3429732 are 8.32%, 9.01%, 8.93%, 8.73%, respectively. The favorable alleles ratio in the population of S18_1766721-G, S09_48870909-T, S19_49517336-G and S18_3429732-G were 37.9%, 34.5%, 87.5%, and 13.6%, respectively (Table 5). And the HST germplasms NPS044, NPS060, NPS151, NPS187 and NPS254 each contained three, three, two, four and one favorable alleles (Table 5). Also, we developed KASP markers for these four SNPs (Table 6). The designed molecular markers effectively differentiate between these two genotypes (Figures 7E–H).

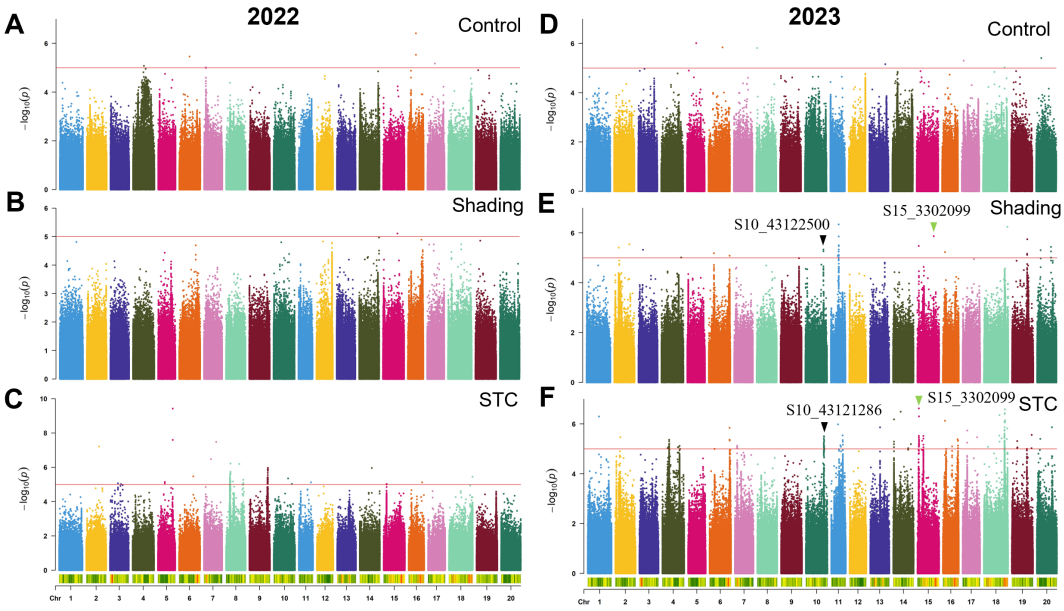


FIGURE 5
GWAS for grain weight with or without shade treatment in 2022 and 2023. **(A–C)**, control, shade treatment and STC of 2022; **(D–F)**, control, shade treatment and STC of 2023, respectively. Red lines represent $-\log_{10}(p) \geq 5.0$.

10.3389/fpls.2024.1479536

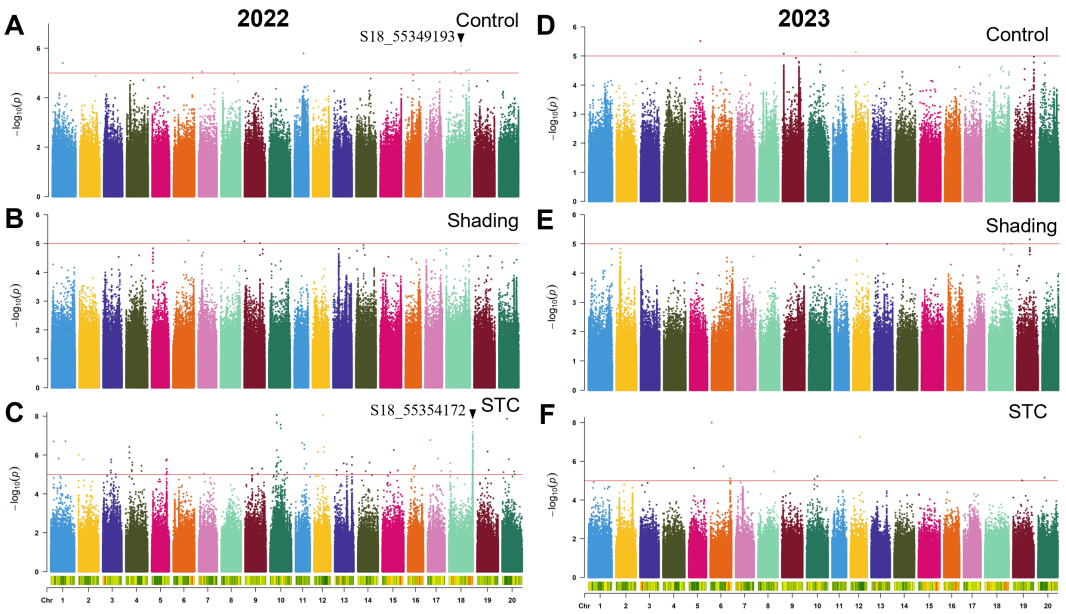


FIGURE 6
GWAS for branch number per plant with or without shade treatment in 2022 and 2023. **(A–C)**, control, shade treatment and STC of 2022; **(D–F)**, control, shade treatment and STC of 2023, respectively. Red lines represent $-\log_{10}(p) \geq 5.0$.

Identification candidate genes for shade tolerance based on GWAS

The LD of this population is 120 kb (Zhang et al., 2021). Therefore, we examined candidate genes within a 120 kb range upstream and downstream of SNPs significantly associated with soybean STC across six traits. Utilizing functional annotation information from the soybean genome, we identified four candidate genes significantly linked to soybean shade tolerance (Table 7). *Glyma.18G024000* associated with S18_1766721, encodes a trichome birefringence-like 33 protein. The gene *Glyma.09G271100* linked to S09_48870909, encodes a protein from the auxin efflux carrier family. *Glyma.19G248900*, associated with S19_49517336, encodes an ethylene response factor 1. Lastly, the gene *Glyma.18G040700*, related to S18_3429732, encodes a MYB domain protein 43.

TABLE 4 GWAS analysis results for six traits associated with shade tolerance.

Env.	Trait	Significant SNP Number	-$\log_{10}(p)$		R^2 (%)	
		-$\log_{10}(p) \geq 5.0$	Max	Min	Max	Min
2022	STC1	24	6.61	5.01	10.87	7.86
	STC2	29	7.15	5.08	12.56	8.42
	STC3	2	6.12	6.12	10.69	10.69
	STC4	13	6.21	5.05	10.94	8.58
	STC5	105	9.28	5.00	18.02	8.74
	STC6	227	8.48	5.01	15.87	8.58
2023	STC1	4	5.25	5.08	9.12	8.78
	STC2	9	6.08	5.06	10.18	8.20
	STC3	2	5.41	5.29	9.44	9.21
	STC4	31	6.61	5.00	12.67	9.15
	STC5	190	6.63	5.00	12.22	8.78
	STC6	97	22.62	5.00	53.76	8.77
Total		733				

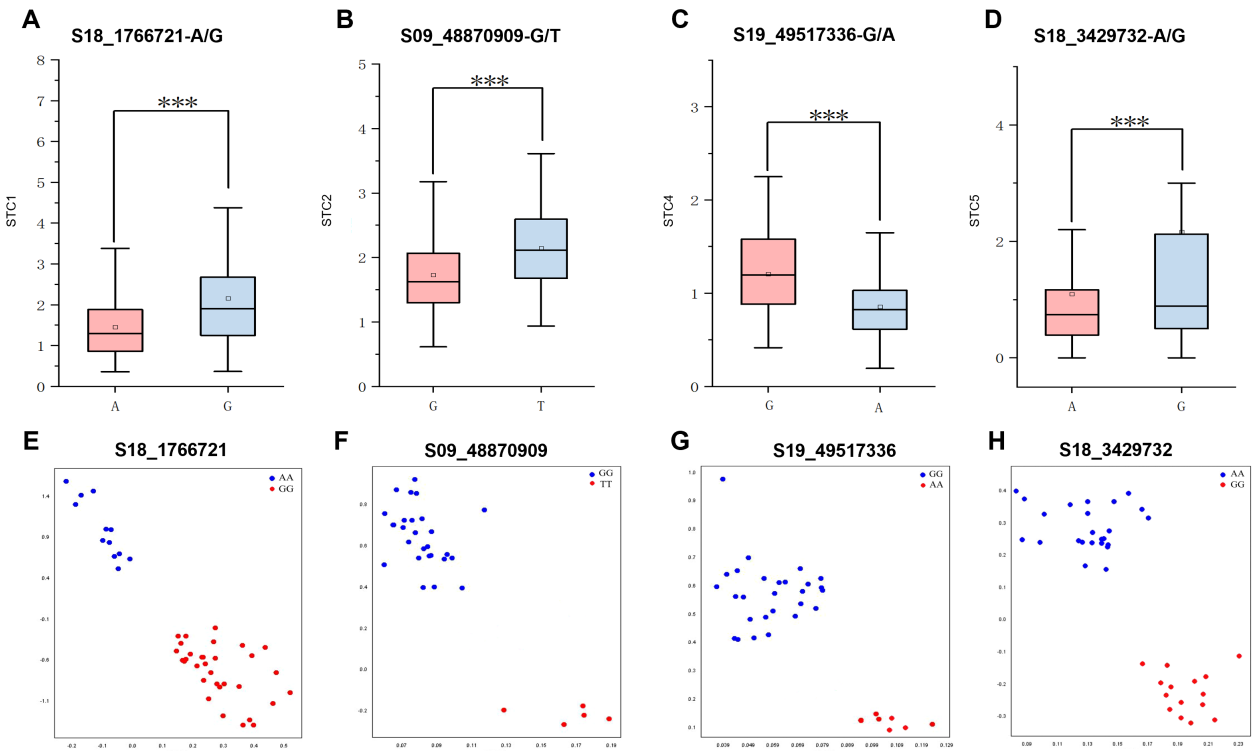


FIGURE 7
Haplotype analysis and genotyping for S18_1766721 **(A, E)**, S09_48870909 **(B, F)**, S19_49517336 **(C, G)** and S18_3429732 **(D, H)**, respectively. *** represents P<0.001.

Discussion

Shade treatment, evaluation and shade-tolerance germplasms

In natural environments, plants are often subjected to shade tolerance. Shade tolerance is essential for soybeans, especially in intercropping or relay cropping systems. When soybeans experience shade stress, their plant height and first pod height will be elongated and the branches, number, grain weight per plant, pod number and nodes number will be reduced, which posed a huge threat to soybean production (Yang et al., 2015; Raza et al., 2020; Table 1).

Previous studies have demonstrated that a 15% reduction in light is considered weak shading, whereas 60% shading often leads to lodging in most varieties, indicating excessive shading. However, at 30% shading, the proportion of lodging varieties and the coefficient of phenotypic variation are sufficient to meet the requirements for shade tolerance identification (Sun et al., 2017; Zhang, 2021). Therefore, this study employed a 30% light reduction to simulate shade treatment. As a result, under 30% shading, all the six traits exhibited more pronounced phenotypic changes, and the result all present normal distribution (Supplementary Table S1; Supplementary Figures S1, S2).

ASFV has been widely utilized for assessing crop resistance to various stressors, including salt, drought and shade tolerance (Zhao et al., 2023). In this study, ASFV was applied to evaluate the shade tolerance of soybean (Figure 2). Five germplasms exhibiting high shade tolerance were identified: NPS044, NPS060, NPS151, NPS187 and NPS254. Notably, NPS044 showed consistent results across two years (Table 3). Chen et al. (2003) measured various parameters such as STC of biological yield during pod setting, plant height, minimum pod height, pod number per plant, grain number per plant, grain weight per plant, and 100 grain weight, and calculated the ASFV of soybean varieties. Similarly, Huang et al. (2012) employed a comprehensive STC across nine indicators, including standard pod

TABLE 5 The number of favorable alleles present in the five high shade-tolerant soybean germplasms.

SNP	Ratio (%)	NPS044	NPS060	NPS151	NPS187	NPS254
S18_1766721-G/A	37.9/62.1	A	G	A	G	A
S09_48870909-T/G	34.5/65.5	T	T	G	T	G
S19_49517336- G/A	87.5/12.5	G	G	G	G	G
S18_3429732-G/A	13.6/86.4	G	A	G	G	A

Ratio represents the allele ratio in the population.

TABLE 6 Primers used for KASP.

SNP	Primer	Primer sequence (5′-3′)
S18_1766721 (STC1)	F1	**GAAGGTGACCAAGTTCATGCT**TAAAAAAAAATGACAATTAG**A**
	F2	**GAAGGTCGGAGTCAACGGATT**TAAAAAAAAATGACAATTAG**G**
	R	TGGCATCCACTCATGAAATCG
S09_48870909 (STC2)	F1	**GAAGGTGACCAAGTTCATGCT**TCATTGATGATAGTATGGTT**G**
	F2	**GAAGGTCGGAGTCAACGGATT**TCATTGATGATAGTATGGTT**T**
	R	GTGTTTCACAACTGCTGGGC
S19_49517336 (STC4)	F1	**GAAGGTGACCAAGTTCATGCT**ATCTAATTTTAATTTACAGT**A**
	F2	**GAAGGTCGGAGTCAACGGATT**ATCTAATTTTAATTTACAGT**T**
	R	ACGAATTGTGTTGGCTGTAACC
S18_3429732 (STC6)	F1	**GAAGGTGACCAAGTTCATGCT**TGTAGAAAACGCGCTTTGTA**A**
	F2	**GAAGGTCGGAGTCAACGGATT**TGTAGAAAACGCGCTTTGTA**G**
	R	TGACAACGACATATGCAAACACAA

Underlines/bold sequences in F1 indicate Field Application Manager (FAM) fluorescent junction sequence and underlines in F2 indicated Hexachlorofluorescein (HEX) fluorescent junction sequence. The bold/underline characters indicate SNPs.

number, standard pod weight, 100 grain weight, plot yield, plant height, main stem node number, number of ingle grain pods number per plant, single plant pod weight per plant, and standard pod length, to determine soybean shade tolerance. Li et al. (2014) developed a mathematical model for evaluating soybean shade tolerance using stepwise regression and identified seven key indicators: main stem node number, branch number, internode length, lodging resistance, pod number per plant, 100-grain weight, and grain weight per plant. Wu et al. (2015) suggest that the rapid identification and prediction of shade tolerance in soybean seedlings can be achieved by measuring leaf dry weight, stomatal conductance, plant height, and maximum fluorescence yield under dark conditions. Tang et al. (2022) used traits such as first pod height, stem node number, pod number per plant, and grain number per pod to evaluate shade tolerance. In this study, six traits - first pod height, plant height, pod number per plant, grain weight per plant, branch number and main stem node number – were measured using STC as the indicator to evaluate shade tolerance in 264 soybean accessions. These traits are reliable for identifying key loci and genes associated with shade tolerance in soybeans.

Shade-tolerance SNPs and candidate genes associated with soybean shade tolerance

A total of 733 SNPs were identified as being associated with the STC of six traits over two years (Figures 3–6; Supplementary Figures S3, S4; Table 4). Due to the significant influence of environmental factors on these traits (Supplementary Figure S1), we didn't co-locate any significant loci between two years. More environments may need to be added.

Based on GWAS, four SNPs S18_1766721, S09_48870909, S19_49517336 and S18_3429732 were selected for further study. Their phenotypic explanation rate ranged from8.32% to 9.01%, which can be used for soybean genome selection breeding. Four candidate genes associate with the four SNPs were identified. *Glyma.18G024000*, associated with S18_1766721, encodes the protein Trichome birefringence-like 33 (TBL33). Members of the TBL family, previously characterized, are localized in the Golgi apparatus and function as polysaccharide O-acetyltransferases catalyzing the O-acetylation of specific cell wall polymers (Stranne et al., 2018; Sinclair et al., 2020; Lunin et al., 2020). TBL proteins have been reported to play roles in biotic (disease, herbivore) and abiotic resistance (salt, drought and freezing) (Xiong et al., 2013; Gao et al., 2017; Sun et al., 2020). Glyma.09G271100 is an auxin efflux carrier family protein, known as PIN -like (PILS) which plays a crucial role in auxin signaling (Bogaert et al., 2022; Feraru et al., 2022; Waidmann et al., 2023). Numerous studies have demonstrated that auxin plays pivotal roles in integrating responses to abiotic stresses such as temperature, water, light and salt and in controlling downstream stress responses (Iglesias et al., 2018; Waadt et al., 2022; Xie et al., 2022; Jing et al., 2023). Organ-specific transcriptome analysis has revealed that shade induces a set of auxin-responsive genes, such as SMALL AUXIN UPREGULATED RNAs (SAURs) and AUXIN/INDOLE-3-ACETIC ACIDs (AUX/IAAs) (Nguyen et al., 2023). In the initial response to shade signals, auxin biosynthesis, transport, and sensitivity are rapidly activated, promoting cell elongation in hypocotyls and other organs (Ma and Li, 2019). *Glyma.19G248900* associated with S19_49517336, encodes an ethylene response factor 1 (ERF1). *GmERF3* has been reported to positively regulates resistance to virus, high salinity and dehydration stresses (Zhang et al., 2009; Liu et al., 2024). Ethylene is known to play a crucial role in mediating plant adaptations to environmental conditions (Waadt et al., 2022). Recent studies have shown that shade stress can induce ethylene biosynthesis, accelerating soybean senescence and hindering nitrogen

TABLE 7 Functional annotation of candidate genes related to shade tolerance in soybean.

Trait	Gene ID	Homologs	Functional annotation
STC1	*Glyma.18G024000*	*AT2G40320*	Trichome birefringence-like 33
STC2	*Glyma.09G271100*	*AT5G01990*	Auxin efflux carrier family protein
STC4	*Glyma.19G248900*	*AT3G23240*	Ethylene response factor 1
STC6	*Glyma.18G040700*	*AT5G16600*	MYB domain protein 43

remobilization (Deng et al., 2024). *ERFs* are significant in enhancing flood tolerance in rice (Xu et al., 2006), where ethylene accumulation in submerged tissues induces the expression of *ERFs* such as SNORKEL1 and SNORKEL2, which are major QTLs associated with deepwater internode elongation (Hattori et al., 2009). *Glyma.18G040700* related to S18_3429732, encodes MYB domain protein 43. This protein plays a critical role in various aspects of plant growth and development, including secondary metabolic regulation, responses to hormones and environmental factors, cell differentiation, organ morphogenesis, and cell cycle regulation (Li et al., 2019a). The homolog AtMYB43 has been reported to be involved in regulating tolerance to cadmium and freezing (Zheng et al., 2022, 2023). In summary, these four genes *Glyma.18G024000*, *Glyma.09G271100*, *Glyma.19G248900*, and *Glyma.18G040700* may be involved in soybean responses to shade tolerance.

KASP markers for soybean shade tolerance

This study developed four KASP markers based on SNPs associated with soybean STC obtained from GWAS. These markers have been successfully used for genotyping (Table 7; Figure 7). Specifically, S18_1766721 is associated with the STC of first pod height, S09_48870909 with the STC of plant height, S19_49517336 with the STC of pod number per plant, and S18_3429732 with the STC of branch number. These markers are valuable tools for identifying shade-tolerant soybean germplasms and can enhance the efficiency and accuracy of selection in molecular marker-assisted breeding. However, KASP markers for the STC of node number and grain weight were not developed. This gap may be due to the influence of multiple factors, suggesting that further efforts and research are needed to identify effective markers for these traits.

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

QJ: Funding acquisition, Methodology, Writing – original draft, Writing – review & editing, Conceptualization. SH: Data curation, Formal analysis, Software, Writing – review & editing. XHL: Methodology, Software, Writing – review & editing. LW: Investigation, Resources, Visualization, Writing – review & editing. QW: Methodology, Supervision, Writing – review & editing. WZ: Formal analysis, Methodology, Software, Writing – review & editing. HZ: Resources, Validation, Writing – review & editing. XQL: Supervision, Validation, Writing – review & editing. XC: Supervision, Validation, Writing – review & editing. XW: Conceptualization, Supervision, Visualization, Writing – review & editing. HC: Conceptualization, Funding acquisition, Resources, Supervision, Visualization, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was financially supported by grants from Jiangsu Key Research and Development Program (BE2022328), Jiangsu Provincial Seed Industry Revitalization “Challenge-and-Select” Project (JBGS (2021)060), Jiangsu Funding Program for Excellent Postdoctoral Talent (2023ZB647). Jiangsu Agricultural Science and Technology Innovation Fund (CX (22) 5002) and National Key Research Development Program of China (2023YFD2000501).

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.2024.1479536/full#supplementary-material

References

Bogaert, K. A., Blomme, J., Beeckman, T., and De Clerck, O. (2022). Auxin's origin: do PILS hold the key? *Trends Plant Sci.* 27, 227–236. doi: 10.1016/j.tplants.2021.09.008

Chen, H., Sun, Z., Yang, S., and Li, C. (2003). Effect of shading on major characters of soybean and preliminary study on the identification method of soybean shade endurance. *Chin. J. Oil Crop Sci.* 25, 78–82.

Deng, J., Huang, X., Chen, J., Vanholme, B., Guo, J., He, Y., et al. (2024). Shade stress triggers ethylene biosynthesis to accelerate soybean senescence and impede nitrogen remobilization. *Plant Physiol. Biochem.* 210, 108658. doi: 10.1016/j.plaphy.2024.108658

Du, Z. (2023). *Identification shade tolerance of soybean germplasm resources and screening of shade tolerance varieties.* (Harbin, China: Master Dissertation of Northeast Agricultural University in China). doi: 10.27010/d.cnki.gdbnu.2023.000975

Echarte, L., Della Maggiora, A., Cerrudo, D., Gonzalez, V. H., Abbate, P., Cerrudo, A., et al. (2011). Yield response to plant density of maize and sunflower intercropped with soybean. *Field Crops Res.* 121, 423–429. doi: 10.1016/j.fcr.2011.01.011

Fankhauser, C., and Batschauer, A. (2016). Shadow on the plant: A strategy to exit. *Cell* 164, 15–17. doi: 10.1016/j.cell.2015.12.043

Feraru, E., Feraru, M. I., Moulinier-Anzola, J., Schwihla, M., Ferreira Da Silva Santos, J., Sun, L., et al. (2022). PILS proteins provide a homeostatic feedback on auxin signaling output. *Development* 149, dev200929. doi: 10.1242/dev.200929

Gao, Y., He, C., Zhang, D., Liu, X., Xu, Z., Tian, Y., et al. (2017). Two trichome birefringence-like proteins mediate xylan acetylation, which is essential for leaf blight resistance in rice. *Plant Physiol.* 173, 470–481. doi: 10.1104/pp.16.01618

Ghosh, P., Tripathi, A., Bandyopadhyay, K., and Manna, M. (2009). Assessment of nutrient competition and nutrient requirement in soybean/sorghum intercropping system. *Eur. J. Agron.* 31, 43–50. doi: 10.1016/j.eja.2009.03.002

Gong, W. Z., Jiang, C. D., Wu, Y. S., Chen, H. H., Liu, W. Y., and Yang, W. Y. (2015). Tolerance vs. avoidance: Two strategies of soybean (Glycine max) seedlings in response to shade in intercropping. *Photosynthetica* 53, 259–268. doi: 10.1007/s11099-015-0103-8

Hattori, Y., Nagai, K., Furukawa, S., Song, X. J., Kawano, R., Sakakibara, H., et al. (2009). The ethylene response factors SNORKEL1 and SNORKEL2 allow rice to adapt to deep water. *Nature* 460, 1026–1030. doi: 10.1038/nature08258

Huang, Q. C., Li, C. Y., Zhao, H. T., Wu, J. M., Zhao, Y. H., Yang, S. Z., et al. (2012). Research of shade-tolerant on vegetable soybean germplasm resources under shading stress. *Southwest China Agric. Sci.* 25, 2212–2217. doi: 10.16213/j.cnki.scjas.2012.06.067

Iglesias, M. J., Sellaro, R., Zurbriggen, M. D., and Casal, J. J. (2018). Multiple links between shade avoidance and auxin networks. *J. Exp. Bot.* 69, 213–228. doi: 10.1093/jxb/erx295

Jia, Q., Zhou, M., Xiong, Y., Wang, J., Xu, D., Zhang, H., et al. (2024). Development of KASP markers assisted with soybean drought tolerance in the germination stage based on GWAS. *Front. Plant Sci.* 15. doi: 10.3389/fpls.2024.1352379

Jiang, A., Liu, J., Gao, W., Ma, R., Zhang, J., Zhang, X., et al. (2023). Transcriptomic and metabolomic analyses reveal the key genes related to shade tolerance in soybean. *Int. J. Mol. Sci.* 24, 14230. doi: 10.3390/ijms241814230

Jing, H., Wilkinson, E. G., Sageman-Furnas, K., and Strader, L. C. (2023). Auxin and abiotic stress responses. *J. Exp. Bot.* 74, 7000–7014. doi: 10.1093/jxb/erad325

Li, J., Han, G., Sun, C., and Sui, N. (2019a). Research advances of MYB transcription factors in plant stress resistance and breeding. *Plant Signal. Behav.* 14, 1613131. doi: 10.1080/15592324.2019.1613131

Li, Y., Jiang, H., Sun, X., Muhammad, A. A., Liu, J., Liu, W., et al. (2019b). Quantitative proteomic analyses identified multiple sugar metabolic proteins in soybean under shade stress. *J. Biochem.* 165, 277–288. doi: 10.1093/jb/mvy103

Li, L., Ljung, K., Breton, G., Schmitz, R. J., Pruneda-Paz, J., Cowing-Zitron, C., et al. (2012). Linking photoreceptor excitation to changes in plant architecture. *Genes Dev.* 26, 785–790. doi: 10.1101/gad.187849.112

Li, S., Sun, Z., Sang, Q., Qin, C., Kong, L., and Huang, X. (2023a). Soybean reduced internode 1 determines internode length and improves grain yield at dense planting. *Nat. Commun.* 14, 7939. doi: 10.1038/s41467-023-42991-z

Li, C., Yao, X., Ju, B., Zhu, M., Wang, H., and Zhang, H. (2014). Analysis of shade-tolerance and determination of shade-tolerance evaluation indicators in different soybean genotypes. *Scientia Agricultura Sin.* 47, 2927–2939.

Li, Z., Zhao, T., Liu, J., Li, H., and Liu, B. (2023b). Shade-induced leaf senescence in plants. *Plants (Basel).* 12, 1550. doi: 10.3390/plants12071550

Liu, F., Xi, M., Liu, T., Wu, X., Ju, L., and Wang, D. (2024). The central role of transcription factors in bridging biotic and abiotic stress responses for plants' resilience. *New crops.* 1, 100005. doi: 10.1016/j.ncrops.2023.11.003

Lunin, V. V., Wang, H. T., Bharadwaj, V. S., Alahuhta, M., Peña, M. J., Yang, J. Y., et al. (2020). Molecular Mechanism of polysaccharide acetylation by the Arabidopsis xylan o-acetyltransferase XOAT1. *Plant Cell.* 32, 2367–2382. doi: 10.1105/tpc.20.00028

Ma, L., and Li, G. (2019). Auxin-dependent cell elongation during the shade avoidance response. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.00914

Martinez-Garcia, J. F., and Rodriguez-Concepcion, M. (2023). Molecular mechanisms of shade tolerance in plants. *New Phytol.* 239, 1190–1202. doi: 10.1111/nph.19047

Nguyen, N. H., Sng, B. J. R., Chin, H. J., Choi, I. K. Y., Yeo, H. C., and Jang, I. C. (2023). HISTONE DEACETYLASE 9 promotes hypocotyl-specific auxin response under shade. *Plant J.* 116, 804–822. doi: 10.1111/tpj.16410

Qin, C., Li, Y. H., Li, D., Zhang, X., Kong, L., Zhou, Y., et al. (2023). PH13 improves soybean shade traits and enhances yield for high-density planting at high latitudes. *Nat. Commun.* 14, 6813. doi: 10.1038/s41467-023-42608-5

Qiu, L. J., and Chang, R. Z. (2006). Descriptors and data standard for soybean (*Glycine* spp.). *Beijing: China Agric. Press*, 55–61.

Raza, M. A., Feng, L. Y., Iqbal, N., Khan, I., Meraj, T. A., Xi, Z. J., et al. (2020). Effects of contrasting shade treatments on the carbon production and antioxidant activities of soybean plants. Funct. *Plant Biol.* 47, 342–354. doi: 10.1071/FP19213

Sinclair, S. A., Gille, S., Pauly, M., and Krämer, U. (2020). Regulation of acetylation of plant cell wall components is complex and responds to external stimuli. *Plant Signal Behav.* 15, 1687185. doi: 10.1080/15592324.2019.1687185

Stranne, M., Ren, Y., Fimognari, L., Birdseye, D., Yan, J., Bardor, M., et al. (2018). TBL10 is required for O-acetylation of pectic rhamnogalacturonan-I in Arabidopsis thaliana. *Plant J.* 96, 772–785. doi: 10.1111/tpj.14067

Su, Y., Yang, H., Wu, Y., Gong, W., Gul, H., Yan, Y., et al. (2023a). Photosynthetic acclimation of shade-grown soybean seedlings to a high-light environment. *Plants (Basel).* 12, 2324. doi: 10.3390/plants12122324

Su, Y., Zhang, Z., He, J., Zeng, W., Cai, Z., Lai, Z., et al. (2023b). Gene-allele system of shade tolerance in southern China soybean germplasm revealed by genome-wide association study using gene-allele sequence as markers. *Theor. Appl. Genet.* 136, 152. doi: 10.1007/s00122-023-04390-2

Sun, A., Yu, B., Zhang, Q., Peng, Y., Yang, J., Sun, Y., et al. (2020). MYC2-activated TRICHOME BIREFRINGENCE-LIKE37 acetylates cell walls and enhances herbivore resistance. *Plant Physiol.* 184, 1083–1096. doi: 10.1104/pp.20.00683

Sun, Z., Zhang, Z., Cai, Z., Zeng, W., Lai, Z., Yang, H., et al. (2017). Establishment of an evaluation system of shade tolerance in soybean and its variation in southern China germplasm population. *Scientia Agricultura Sin.* 50, 792–801.

Waadt, R., Seller, C. A., Hsu, P. K., Takahashi, Y., Munemasa, S., and Schroeder, J. I. (2022). Plant hormone regulation of abiotic stress responses. *Nat. Rev. Mol. Cell Biol.* 23, 680–694. doi: 10.1038/s41580-022-00479-6

Waidmann, S., Béziat, C., Ferreira Da Silva Santos, J., Feraru, E., Feraru, M. I., Sun, L., et al. (2023). Endoplasmic reticulum stress controls PIN-LIKES abundance and thereby growth adaptation. *Proc. Natl. Acad. Sci. U. S. A.* 120, e2218865120. doi: 10.1073/pnas.2218865120

Wang, Z., Wang, Y., Chen, M., Wu, J., Yang, H., Yang, W., et al. (2023). Effects of different shade degrees and light recovery on photosynthetic physiology of different soybean varieties. *J. Sichuan Agric. Univ.* 41, 765–772. doi: 10.16036/j.issn.1000-2650.202303225

Wu, X., Liang, H., Yang, F., Liu, W., She, Y., and Yang, W. (2015). Comprehensive evaluation and screening identification indexes of shade tolerance at seedling in soybean. *Scientia Agricultura Sin.* 48, 2497–2507.

Wu, Y. S., Yang, F., Gong, W. Z., Ahmed, S., Fan, Y. F., Wu, X. L., et al. (2017). Shade adaptive response and yield analysis of different soybean genotypes in relay intercropping systems. *J. Integr. Agric.* 16, 1331–1340. doi: 10.1016/S2095-3119(16)61525-3

Xie, X., Cheng, H., Hou, C., and Ren, M. (2022). Integration of light and auxin signaling in shade plants: from mechanisms to opportunities in urban agriculture. *Int. J. Mol. Sci.* 23, 3422. doi: 10.3390/ijms23073422

Xiong, G., Cheng, K., and Pauly, M. (2013). Xylan O-acetylation impacts xylem development and enzymatic recalcitrance as indicated by the Arabidopsis mutant tbl29. *Mol. Plant* 6, 1373–1375. doi: 10.1093/mp/sst014

Xu, K., Xu, X., Fukao, T., Canlas, P., Maghirang-Rodriguez, R., Heuer, S., et al. (2006). *Sub1A* is an ethylene-response-factor-like gene that confers submergence tolerance to rice. *Nature* 442, 705–708. doi: 10.1038/nature04920

Yang, F., Wang, X. C., Liao, D. P., Lu, F. Z., Gao, R. C., Liu, W. G., et al. (2015). Yield response to different planting geometries in maize–soybean relay strip intercropping systems. *Agron. J.* 107, 296–304. doi: 10.2134/agronj14.0263

Zhang, Z. (2021). *Geographic-seasonal differentiation and genetic analysis of shade-tolerance related traits in southern China soybean germplasm population*. (Nanjing, China: Doctoral Dissertation of Nanjing Agricultural University in China). doi: 10.27244/d.cnki.gnjnu.2019.000182

Zhang, G., Chen, M., Li, L., Xu, Z., Chen, X., Guo, J., et al. (2009). Overexpression of the soybean *GmERF3* gene, an AP2/ERF type transcription factor for increased tolerances to salt, drought, and diseases in transgenic tobacco. *J. Exp. Bot.* 60, 3781–3796. doi: 10.1093/jxb/erp214

Zhang, W., Xu, W., Zhang, H., Liu, X., Cui, X., Li, S., et al. (2021). Comparative selective signature analysis and high-resolution GWAS reveal a new candidate gene controlling seed weight in soybean. *Theor. Appl. Genet.* 134, 1329–1341. doi: 10.1007/s00122-021-03774-6

Zhao, Z., Fu, M., Li, S., Wang, Y., Yu, X., Zhang, H., et al. (2023). Shading-tolerant spring soybean: Gentypes screening and identification of shade tolerance indexes. *Chin. Agric. Sci. Bull.* 39, 25–32.

Zheng, P., Cao, L., Zhang, C., Fang, X., Wang, L., and Miao, M. (2023). The transcription factor MYB43 antagonizes with ICE1 to regulate freezing tolerance in Arabidopsis. *New Phytol.* 238, 2440–2459. doi: 10.1111/nph.18882

Zheng, P., Cao, L., Zhang, C., Pan, W., Wang, W., Yu, X., et al. (2022). MYB43 as a novel substrate for CRL4PRL1 E3 ligases negatively regulates cadmium tolerance through transcriptional inhibition of HMAs in Arabidopsis. *New Phytol.* 234, 884–901. doi: 10.1111/nph.18020

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 18 November 2024
DOI 10.3389/fpls.2024.1494551

OPEN ACCESS

EDITED BY
Huatao Chen,
Jiangsu Academy of Agricultural Sciences (JAAS), China

REVIEWED BY
Zhandong Cai,
South China Agricultural University, China
Jiajia Li,
Anhui Agricultural University, China

*CORRESPONDENCE
Long Yan
dragonyan1979@163.com
Peijun Tao
taopeijun@sina.com
Ainong Shi
Ainong_shi@hotmail.com

†These authors have contributed equally to this work

RECEIVED 11 September 2024
ACCEPTED 07 October 2024
PUBLISHED 18 November 2024

CITATION
Xu R, Yang Q, Liu Z, Shi X, Wu X, Chen Y, Du X, Gao Q, He D, Shi A, Tao P and Yan L (2024) Genome-wide association analysis and genomic prediction of salt tolerance trait in soybean germplasm.
Front. Plant Sci. 15:1494551.
doi: 10.3389/fpls.2024.1494551

Genome-wide association analysis and genomic prediction of salt tolerance trait in soybean germplasm

Rongqing Xu[1,2†], Qing Yang[1†], Zhi Liu[1], Xiaolei Shi[1], Xintong Wu[1], Yuehan Chen[1], Xinyu Du[1,2], Qiqi Gao[1], Di He[1], Ainong Shi[3*], Peijun Tao[2*] and Long Yan[1*]

[1]Hebei Laboratory of Crop Genetics and Breeding, National Soybean Improvement Center Shijiazhuang Sub-Center, Huang-Huai-Hai Key Laboratory of Biology and Genetic Improvement of Soybean, Ministry of Agriculture and Rural Affairs, Institute of Cereal and Oil Crops, Hebei Academy of Agricultural and Forestry Sciences, Shijiazhuang, China, [2]College of Agronomy, Hebei Agricultural University, Baoding, China, [3]Department of Horticulture, University of Arkansas, Fayetteville, AR, United States

Introduction: Soybean is an important protein and oil crop, and improving yield has traditionally been a major breeding goal. However, salt stress is an important abiotic factor that can severely impair soybean yield by disrupting metabolic processes, inhibiting photosynthesis, and hindering plant growth, ultimately leading to a decrease in productivity.

Methods: This study utilized phenotypic and genotypic data from 563 soybean germplasms sourced from over 20 countries. Employing four distinct models—we performed a genome-wide association study (GWAS) using four models, including MLM, MLMM, FarmCPU, and BLINK in GAPIT 3, we conducted a Genome-Wide Association Study (GWAS) to identify single nucleotide polymorphism (SNP) associated with salt tolerance in soybean. Subsequently, these identified SNP were further analyzed for candidate gene discovery. Using 34,181 SNPs for genomic prediction (GP) to assess prediction accuracy.

Results: Our study identified 10 SNPs significantly associated with salt tolerance, located on chromosomes 1, 2, 3, 7, and 16. And we identified 11 genes within a 5 kb window upstream and downstream of the QTLs on chromosomes 1, 3, and 16. Utilizing the GWAS-derived SNP marker sets for genomic prediction (GP) yielded r-values greater than 0.35, indicating a higher level of accuracy. This suggests that genomic selection for salt tolerance is feasible.

Discussion: The 10 identified SNP markers and candidate genes in this study provide a valuable reference for screening and developing salt-tolerant soybean germplasm resources.

KEYWORDS

soybean, salt stress, genome-wide association study, genomic prediction, germplasm

Introduction

Soybean is widely cultivated in the world, it is an important food and economic crop, ranking sixth in global food crop production (Du et al., 2023; Lu et al., 2020; Wang et al., 2023). As the most significant legumes globally, soybean is rich in protein, oil, isoflavones, and dietary fiber, providing high nutritional value (Graham and Vance, 2003). It also offer health benefits, including enhanced human immunity, prevention of cardiovascular diseases, and potential anti-aging effects. With improving living standards, the demand for soybean products has steadily increased. However, the average global yield of soybeans, approximately 2.5 to 3 tons per hectare, is insufficient to meet this growing demand. As a result, increasing soybean yield has become a priority for breeders.

Soil salinity, a major abiotic stress factor, significantly inhibits seed germination, growth, and nodule formation in soybeans (Ondrasek et al., 2011; Singleton and Bohlool, 1984). Data from the Food and Agriculture Organization of the United Nations and the United Nations Environment Program reveal that over 950,000 square kilometers of land worldwide have been degraded to saline-alkali conditions, accounting for more than 8% of the global land area (Beecher, 1994). Therefore, breeding salt-tolerant soybean varieties is essential for enhancing soybean production.

There has been extensive research on the genes and quantitative trait loci (QTLs) associated with salt tolerance in soybeans. To date, 1,536 QTLs related to salt tolerance have been identified, primarily distributed across chromosomes 2, 3, 6, 8, 9, 12, 13, 14, and 17. Utilizing 196 soybean landraces and 184 families, Kan et al. (2016) identified 22 SSR markers tightly linked to salt tolerance during germination, as well as 11 QTL loci on chromosomes 2, 7, 8, 10, 17, and 18. Similarly, Chen et al. (2008) conducted QTL mapping for salt tolerance at the seedling stage in soybeans using a RIL population of 184 lines, identified eight QTL loci on chromosomes 2, 3, 7, 9, 11, 14, and 18. In another separate study, Huang (2013) conducted visual leaf scorch scoring under salt stress and identified 62 SNP markers on chromosomes 2, 3, 5, 6, 8, and 18 that were significantly associated with salt tolerance.

Genome-Wide Association Study (GWAS) has become a favored method for studying the association between complex traits and genetic variations across the genome due to its high efficiency and shorter time required for constructing populations. Kan et al. (2016) conducted a GWAS of four salt tolerance indices using 191 soybean germplasms genotyped with 1,142 SNPs and identified eight SNP markers and five candidate genes associated with salt stress. Zhang et al. (2019) performed salt inhibition seed germination experiments on 211 cultivated soybean germplasms and conducted a GWAS of four salt tolerance indices with 207,608 SNPs from the NJAU 355 K SoySNP database (CMLM model). They detected 92 trait markers on chromosomes 1, 8, 11, 13, 14, 15, 16, 18, and 19. Further integration of QTL mapping results from 184 RILs and gene expression analysis identified a candidate gene, *Glyma.08g102000*, which belongs to the cation diffusion facilitator (CDF) family, for salt tolerance. Transgenic verification confirmed the gene's role in regulating salt tolerance. In another study, Patil et al. (2016) measured chloride concentration and chlorophyll content in the leaves of 106 soybean lines at the V2 stage and conducted association analysis with 37,000 SNP markers from the SoySNP50K database. They identified 30 SNPs on chromosome 3 significantly associated with chlorophyll content and leaf wilting degree. Zeng et al. (2017) used 283 soybeangermplasm collected worldwide and measured chloride concentration and chlorophyll content in leaves at the V1 stage as salt tolerance indicators. They conducted a GWAS using 33,000 SNP markers from the SoySNP50K dataset (Song et al., 2015) and identified 45 SNPs on chromosomes 2, 3, 7, 8, 10, 13, 14, 16 and 20, and 31 SNPs on chromosome 3 significantly associated with salt tolerance.

Genomic Prediction (GP) can significantly accelerate breeding process. Molecular breeding of crops, particularly through Genomic Selection (GS), is an approach in crop genetic improvement. Compared to marker-assisted selection (MAS), GS can capture genetic effect, shorten breeding cycle and improve breeding efficiency (Hickey et al., 2017). Duhnen et al. (2017) reported an average prediction accuracy of 0.39 for soybean yield, and as high as 0.80 in some population (Jarquín et al., 2014). GP plays a crucial role in GS, allowing researchers to predict crop traits across various environments (Keller et al., 2020; Shikha et al., 2017).

Currently, there are 563 soybean germplasms with salt phenotypic data available in the United States Department of Agriculture (USDA) Germplasm Resources Information Network (GRIN) database (https://npgsweb.ars-grin.gov/gringlobal/descriptordetail?id=51054). Of these, 563 germplasms also have SNP genotypic data in SoyBase (https://www.soybase.org/snps/; Song et al., 2015). The purpose of this study is to utilize these phenotypic and genotypic data to conduct a GWAS to identify SNP markers associated with salt tolerance in soybean. Additionally, GP was performed to evaluate its potential application in selecting salt-tolerant lines for soybean breeding programs.

Materials and methods

Plant material

The 563 soybean germplasms from the USDA Germplasm Collection were used for this study. These germplasms were originally collected from 26 countries, including Japan (159 germplasms), China (86), India (59), South Korea (46), North Korea (28), South Africa (26), Nepal (20), Indonesia (16), United States (16), Suriname (13), Thailand (10), and 15 other countries (39), plus 45 germplasms of unknown origin (Supplementary Tables 1A, B).

Phenotyping

The phenotypic data for salt tolerance reactions in 563 soybean germplasms were downloaded from the USDA GRIN website: https://npgsweb.ars-grin.gov/gringlobal/descriptordetail?id=51054. The experiments were conducted in Illinois, United States by Randy Nelson at the USDA Soybean Collection in Urbana, IL. The salt reaction was scored as tolerant (T) and susceptible (S) for each accession. '1' indicating tolerance and '9' indicating susceptibility were used to perform GWAS to identify SNP markers associated with salt tolerance in this study (Supplementary Table 1A).

Genotyping

The germplasms were genotyped using Soy50K SNP Infinium Chips (Song et al., 2013). A total of 42,292 SNPs across 563 soybean germplasms were downloaded from SoyBase at https://soybase.org/snps/download.php (Song et al., 2015). For GWAS, 34,181 SNPs were selected after excluding those with more than 5% missing data, heterozygosity greater than 5%, and a minor allele frequency (MAF) less than 5%. These SNPs were distributed across all 20 chromosomes of the soybean genome (Supplementary Figure 1).

Principal component analysis and genetic diversity

In this study, 34,181 SNPs were included in the principal component analysis (PCA) and genetic diversity analysis. PCA and genetic diversity were analyzed using GAPIT 3 (Wang and Zhang, 2021), with PCA components set from 2 to 10 and NJ tree settings from 2 to 10. Phylogenetic trees were drawn using the neighbor-joining (NJ) method in GAPIT 3. Genetic diversity was assessed for all 563 tested germplasms and their salt tolerance using (1) 34,181 SNPs in GAPIT 3, and (2) 10,000 randomly selected SNPs in MEGA 7 (Kumar et al., 2016). The phylogenetic trees were drawn using MEGA 7 based on the Maximum Likelihood method with the parameters described in Shi et al. (2016, 2017).

Association analysis

GWAS was conducted using various models, including Bayesian-information and Linkage-disequilibrium Iteratively Nested Keyway (BLINK), Fixed and Random Model Circulating Probability Unification (FarmCPU), Generalized Linear Model (GLM), and Mixed Linear Model (MLM) in GAPIT 3 (Wang and Zhang, 2021). The analysis was performed on a panel of 563 germplasms using 34,181 SNPs. Multiple GAPIT models were utilized to identify robust and consistent SNP markers associated with salt tolerance in soybean. The significance threshold for germplasms was determined using Bonferroni correction of *P*-values with an $\alpha = 0.05$ (0.05/SNP number). An LOD (logarithm of odds) value of 5.83 [Here, we use LOD instead of –log(*P*-value)] was used as the significance threshold based on the 34,181 SNPs.

Candidate gene prediction

Candidate genes associated with salt tolerance were sought within a 5 kb vicinity on both sides of the significant SNPs, following the methodology outlined by Zhang HY, et al., 2016. The candidate genes were extracted from the reference annotation of the soybean genome assembly, Wm82.a2.v1, available at https://phytozome-next.jgi.doe.gov/info/Gmax_Wm82_a2_v1.

Genomic prediction for genomic selection of salt tolerance

In this investigation, ridge regression best linear unbiased prediction (RR-BLUP) from the rrBLUP package (Endelman, 2011) and Bayesian models, including Bayes A (BA), Bayes B (BB), Bayes LASSO (BL), and Bayes ridge regression (BRR), implemented in the BGLR package, were employed for predicting genomic estimated breeding values (GEBV) in GP. The analysis was carried out using R software version 4.3.1 (https://www.r-project.org/). Additionally, GEBV prediction was conducted using genomic best linear unbiased prediction (gBLUP), composite BLUP (cBLUP), marker-assisted BLUP (maBLUP), and settlement of MLM under progressively exclusive relationship (SUPER) BLUP (sBLUP) methods, implemented in the GAPIT package. The effectiveness of genomic prediction using these approaches has been documented in prior research studies (Shi et al., 2021; 2022; Jarquín et al., 2014; 2017; Zhang JP, et al., 2016).

Genomic prediction (r-value) for salt tolerance was conducted across various soybean panels and scenarios. Firstly, GP was estimated using a training set to predict salt tolerance in the panel of 563 soybean germplasms. Predictions were estimated using four models: maBLUP, gBLUP, cBLUP, and sBLUP, utilizing all 34,181 SNPs in GAPIT3. Secondly, GP was executed using ten SNP sets: eight randomly selected SNP sets ranging from 10 to 10,000 SNPs, plus a GWAS-derived SNP marker set containing 10 markers (m10). These predictions were estimated using five GP models: BA, BB, BL, BRR, and rrBLUP. The prediction accuracy for salt tolerance was assessed using the average Pearson's correlation coefficient (r) between the GEBVs and observed values in the validation set. Training and validation sets were randomly created 100 times, and the r-value was estimated for each iteration. The average r-value across the 100 iterations was then calculated for salt tolerance. In the GP scenarios, a higher r-value indicates greater prediction accuracy and better selection efficiency in GS, reflecting the reliability of the GP for salt tolerance.

Results

Evaluation of salt tolerance

The 563 soybean germplasms were divided into two groups, where 150 germplasms were salt tolerant and 413 were susceptible (Figure 1; Supplementary Tables 1A, C). These 150 salt-tolerant germplasms can be used as parents in soybean breeding programs to develop salt-tolerant lines.

Genome-wide association study

Using GAPIT 3, the 563 soybean germplasms were divided into four distinct clusters (subpopulations), labeled Q1 to Q4, based on the analysis of 34,181 SNPs (Figure 2; Supplementary Figure 2). The clustering was derived from the following analyses: (1) a 3D graphical

10.3389/fpls.2024.1494551

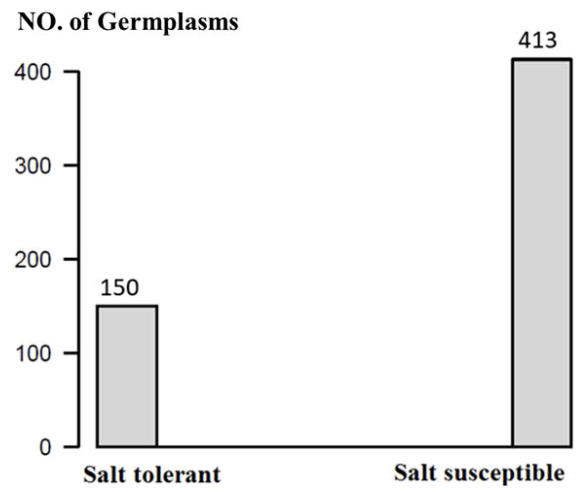


FIGURE 1
The distribution of salt tolerance reaction in 563 soybean germplasms.

plot of the principal component analysis (PCA) (Supplementary Figure 2, left), (2) a PCA eigenvalue plot (Supplementary Figure 2, right), and (3) phylogenetic trees constructed using the neighbor-joining (NJ) method (Figure 2A, ring – left and Figure 2B, no-root - right). Additionally, the kinship plot confirmed the existence of these four groups among the 563 germplasms (Supplementary Figure 3). Each was assigned to one of the four clusters (Q1 to Q4) (Supplementary Table 1A), and the resulting Q-matrix with four clusters was subsequently applied to the GWAS analysis.

Based on the analysis using four models (GLM, MLM, FarmCPU, and BLINK) in GAPIT 3, the multiple QQ plots showed a significant deviation from the expected distribution (Figure 3, right half; Supplementary Figure 4, right), indicating the presence of SNPs associated with salt tolerance. The multiple Manhattan plots, covering all 34,181 tested SNPs, revealed several SNPs with LOD values greater than 5.83, primarily located on chromosomes 1, 2, 3, 7, and 16, suggesting the SNPs were associated with salt tolerance in the panel (Figure 3, left half; Supplementary Figure 4, left).

Ten SNPs with LOD values greater than 5.83 were detected by at least one model (GLM or FarmCPU) in GAPIT 3 for salt tolerance (Table 1). Three SNP markers associated with salt tolerance, Gm01_4306329_ss715579436, Gm01_4312808_ss715579441, and Gm01_4336306_ss715579451, are in a region from 4,306,329 bp to 4,336,306 bp, with an interval of 30 kb on chromosome 1. The LODs of these SNPs were greater than 6.5 in the GLM model (Table 1). The other three SNP markers associated with salt tolerance, Gm02_36878905_ss715582154, Gm02_36940321_ss715582156, and Gm02_36991983_ss715582157, are located in a region from 36,878,905 bp to 36,991,983 bp on chromosome 2. These SNPs had an LOD greater than 6.7 in the GLM model and greater than 4.1 in a *t*-test (Table 1), indicating the presence of a QTL in this region. The two SNP markers, Gm03_43213208_ss715586397 and Gm03_43220331_ss715586399, are located at 43,213,208 bp and 43,220,331 bp, respectively, on chromosome 3. These SNPs showed an LOD greater than 5.83 in either the GLM or FarmCPU model and greater than 8.9 in t-tests (Table 2), suggesting a QTL in this region of chromosome 3 for salt tolerance. The SNP marker Gm07_41776639_ss715598058, located at 41,776,639 bp on chromosome 7, is associated with salt tolerance with an LOD of 6.44 in the GLM model and 5.94 in the t-test (Table 2). The

TABLE 1 List of ten SNPs with LOD (-log(P-value)) greater than 5.83 detected by one or more models (FarmCPU or GLM) in GAPIT 3, along with t-test results for salt tolerance.

SNP	Chr	Pos	MAF %	LOD [-log (P-value)]	Model	Lod_ (*t*-test)	Beneficial _allele	unbeneficial _allele	Linked gene (0-4kb)
Gm01_4306329_ss715579436	1	4306329.00	22.29	6.82	GLM	7.26	A	G	Glyma.01G039600 Glyma.01G039700
Gm01_4312808_ss715579441	1	4312808.00	23.89	6.81	GLM	7.30	A	G	Glyma.01G039700 Glyma.01G039800
Gm01_4336306_ss715579451	1	4336306.00	22.38	6.60	GLM	6.58	G	A	Glyma.01G040000
Gm02_36878905_ss715582154	2	36878905.00	48.49	7.45	GLM	4.11	G	A	Glyma.02G195400
Gm02_36940321_ss715582156	2	36940321.00	34.99	6.73	GLM	4.42	A	G	
Gm02_36991983_ss715582157	2	36991983.00	34.81	6.87	GLM	4.52	A	G	
Gm03_43213208_ss715586397	3	43213208.00	45.74	5.85	GLM	8.97	T	G	Glyma.03G230400 Glyma.03G230500 Glyma.03G230600
Gm03_43220331_ss715586399	3	43220331.00	44.85	6.11	GLM	9.33	C	T	Glyma.03G230600 Glyma.03G230700
				6.03	FarmCPU	9.33			
Gm07_41776639_ss715598058	7	41776639.00	43.78	6.44	GLM	5.94	T	C	
Gm16_7722217_ss715625494	16	7722217.00	11.55	5.95	FarmCPU	2.75	T	C	Glyma.16G076200 Glyma.16G076300

Each SNP is located on a chromosome, and its position is based on the soybean genome reference Wm82.a2 and its linked genes within less than 4 kb.
Chr, chromosome; Pos, position; MAF, minor allele frequency; LOD, Logarithm of the Odds.

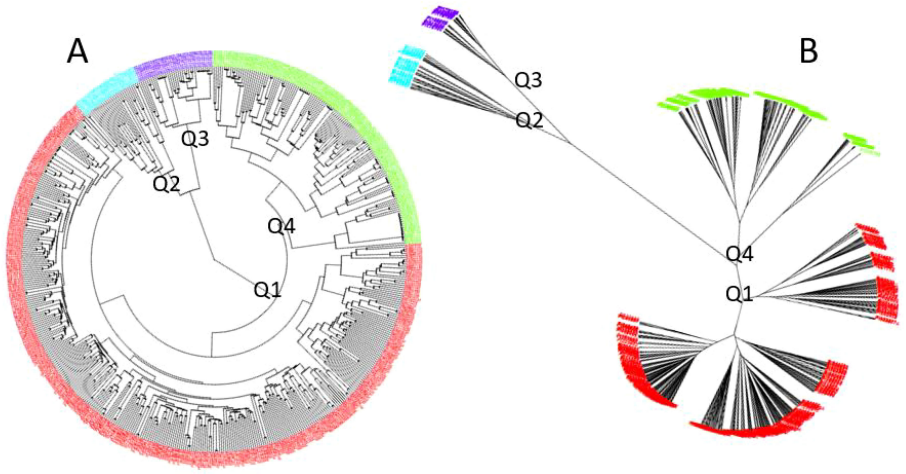


FIGURE 2
Population genetic diversity analysis in the association panel consisting of 563 USDA soybean germplasms: phylogenetic trees [**(A)**. fan and **(B)**. unrooted] drawn using the neighbor-joining (NJ) method in four sub-populations (Q1-Q4) by GAPIT3.

SNP marker Gm16_7722217_ss715625494, located at 7,722,217 bp on chromosome 16, is associated with salt tolerance with an LOD of 5.95 in the GLM model and 2.75 in the t-test (Table 2).

Candidate genes for salt tolerance

Eleven genes are located within 5 kb upstream and downstream of 7 of the 10 SNP markers associated with salt tolerance (Table 1; Supplementary Table 2). Gene information is based on the soybean reference genome *Glycine max* Wm82.a2.v1 (https://phytozome-next.jgi.doe.gov/info/Gmax_Wm82_a2_v1).

The four genes, *Glyma.01G039600* (leucine-rich repeat receptor-like protein kinase family protein), *Glyma.01G039700* (Vps51/Vps67 family protein, components of vesicular transport), *Glyma.01G039800* (galactosyltransferase family protein), and *Glyma.01G040000* (glutathione S-transferase TAU 18), are physically close to the three SNP markers—Gm01_4306329_ss715579436, Gm01_4312808_ss715579441, and Gm01_4336306_ss715579451— (Tables 1, 2).

Similarly, candidate genes *Glyma.02G195400* (syntaxin of plants 121), located at 36,872,648 bp to 36,875,320 bp on chromosome 2, *Glyma.03G230400* (invertase H), *Glyma.03G230500* (plus-3 domain-containing protein), *Glyma.03G230600* (protein of unknown function, DUF538), and *Glyma.03G230700* (importin alpha isoform 4) on chromosome 3 and *Glyma.16G076200* (pyrimidin 4) and *Glyma.16G076300* (long-chain fatty alcohol dehydrogenase family protein on chromosome 16 were within physically close to the significant SNPs on those chromosomes (Tables 1, 2).

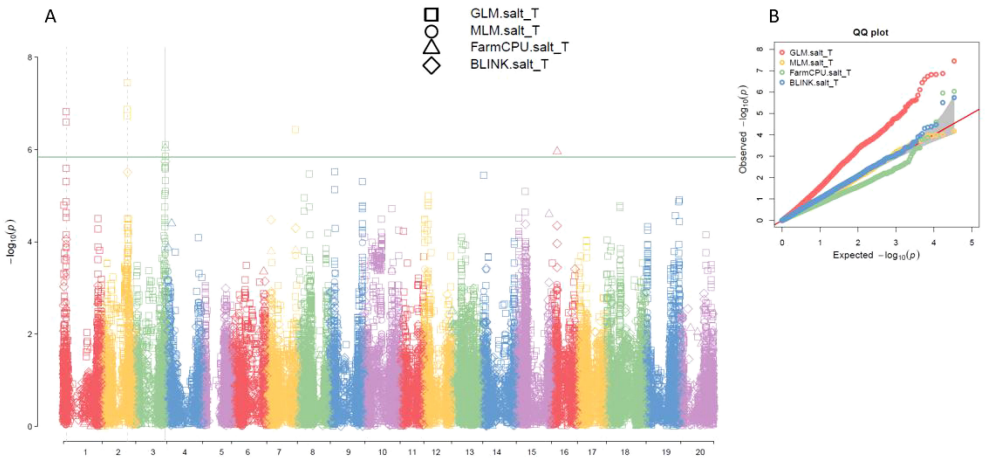


FIGURE 3
Multiple Manhattan plot **(A)** and QQ plot **(B)** of SNP significant level for salt tolerance among GLM, MLM, FarmCPU, and BLINK models in GAPIT3 in an association panel consisting of 563 germplasms. The Manhattan plot (left) illustrates soybean 20 chromosomes on the x-axis and LOD (-log(P-value)) values on the y-axis. The QQ plot (right) displays LOD (-log(P-value)) values on the x-axis and expected LOD (-log(P-value)) values on the y-axis.

TABLE 2 Genomic prediction (r-value) of salt tolerance using nine SNP sets: eight randomly selected SNP sets ranging from 10 to 10,000 SNPs (r10 to r10000), plus the GWAS-derived SNP marker sets (10 markers - m10).

GP Model	r-value						SE of r-value					
	rrBLUP	BA	BB	BL	BRR	SNP.set Mean	rrBLUP	BA	BB	BL	BRR	SNP.set Mean
r10	0.20	0.20	0.14	0.20	0.19	0.19	0.067	0.072	0.088	0.079	0.083	0.078
r100	0.17	0.24	0.22	0.24	0.22	0.22	0.083	0.076	0.075	0.084	0.070	0.078
r200	0.24	0.24	0.25	0.26	0.24	0.25	0.070	0.087	0.086	0.078	0.081	0.080
r500	0.28	0.29	0.29	0.30	0.29	0.29	0.081	0.090	0.084	0.071	0.092	0.084
r1000	0.30	0.31	0.29	0.30	0.31	0.30	0.074	0.077	0.067	0.076	0.075	0.074
r2000	0.33	0.30	0.29	0.31	0.31	0.31	0.085	0.087	0.087	0.077	0.076	0.083
r5000	0.32	0.28	0.30	0.29	0.31	0.30	0.072	0.077	0.085	0.081	0.078	0.079
r10000	0.30	0.27	0.28	0.27	0.30	0.28	0.083	0.084	0.072	0.075	0.080	0.079
m10	0.36	0.38	0.38	0.40	0.39	0.38	0.082	0.079	0.088	0.073	0.077	0.080
GP.model mean	0.28	0.28	0.27	0.29	0.28	0.28	0.077	0.081	0.081	0.077	0.079	0.079

Predictions were estimated using five genomic prediction (GP) models: rrBLUP, BA, BB, BL, and BRR. The standardized errors of the r-values (SE) are also listed.
r-value, the genomic prediction r value for salt tolerance; SE of r-value, the standard error of the genomic prediction r value for salt tolerance; rrBLUP, ridge regression best linear unbiased prediction; BA, Bayes A; BB, Bayes B; BL, Bayes LASSO; BRR, Bayes ridge regression; SNP.set Mean, the average values of the five different genomic prediction models; r10 - r10000, the SNP sets randomly selected in quantities ranging from 10 to 10,000.

Genomic prediction

GP in the reference: The GP analysis yielded moderate to high r-values of 0.46, 0.60, and 0.44 for the maBLUP, gBLUP, and sBLUP models, respectively. These estimates were obtained using a training set to predict salt tolerance in a panel of 563 soybean germplasms genotyped with 34,181 SNPs in (Supplementary Figure 5). These results indicate that GS is effective for salt tolerance selection.

GP in cross-prediction using randomly selected SNP markers: GP using randomly selected SNP markers for cross-prediction yielded the following average r-values: 0.19 (ranging from 0.14 to 0.20) for the 10-SNP set (r10); 0.22 (ranging from 0.17 to 0.24) for the 100-SNP set (r100); 0.25 (ranging from 0.24 to 0.26) for the 200-SNP set (r200); 0.29 (ranging from 0.28 to 0.30) for the 500-SNP set (r500); 0.30 (ranging from 0.29 to 0.31) for the 1,000-SNP set (r1000); 0.31 (ranging from 0.29 to 0.33) for the 2,000-SNP set (r2000); 0.30 (ranging from 0.28 to 0.32) for the 5,000-SNP set (r5000); and 0.28 (ranging from 0.27 to 0.30) for the 10,000-SNP set (r10000) (Table 2; Figure 4). These results demonstrate that the r-value increased with the number of randomly selected SNPs, with an average r-value rising from 0.19 in the 10-SNP set to 0.30 in the 10,000-SNP set. This suggests that a randomly selected SNP set consisting of at least 1,000 SNPs (r = 0.30) should be used in GS for selecting salt tolerance.

GWAS-derived SNP marker set: The average r-value was 0.38, ranging from 0.36 to 0.40, for the 10-marker set (m10) (Table 2; Figure 4). These results indicate that the r-value was moderately high, exceeding 0.35 and surpassing those of SNP sets randomly selected from 10 SNPs to 10,000 SNPs. This suggests that GWAS-derived SNP markers can be effectively used for GP and for selecting salt tolerance in soybean breeding through MAS and GS.

GP Model: All five GP models—BA, BB, BL, BRR, and rrBLUP—exhibited similar r-values, indicating that each model is effective for selecting salt tolerance in GS.

Genetic diversity and utilization of the salt tolerant germplasms

The phylogenetic analysis showed that the 150 salt-tolerant germplasms were distributed throughout the tree of 563 germplasms and did not separate into distinct groups of susceptible and tolerant germplasms (Supplementary Figure 6). This indicates that the 150 salt-tolerant germplasms have broad genetic backgrounds. Further analysis revealed that these 150 salt-tolerant germplasms can be divided into three distinct groups (Supplementary Figures 7, 8), confirming that they possess different genetic backgrounds.

Among the 150 salt-tolerant germplasms, six countries contributed a total of 104 germplasms: Japan (46), China (31), India (12), Suriname (7), Indonesia (6), and Nepal (5) (Supplementary Table 1C). Phylogenetic analysis of these 104 germplasms showed that germplasms from the same country generally clustered together (Supplementary Figure 9), suggesting that germplasms from the same country share similarities in genetic backgrounds. Specifically, germplasms from Nepal are closer to those from Japan, followed by China; germplasms from Indonesia are closer to those from Suriname; while germplasms from India are more distinct (Figure 5). This clustering suggests that geographic factors influence the distribution of salt-tolerant germplasms.

Discussion

Soybeans are a crucial source of plant protein, accounting for over 60% of daily plant protein consumption (Sedivy et al., 2017). With the increasing global demand for food, soybean production must be enhanced to meet the rising need for plant protein (Lu et al., 2021). However, soybean yields are highly susceptible to

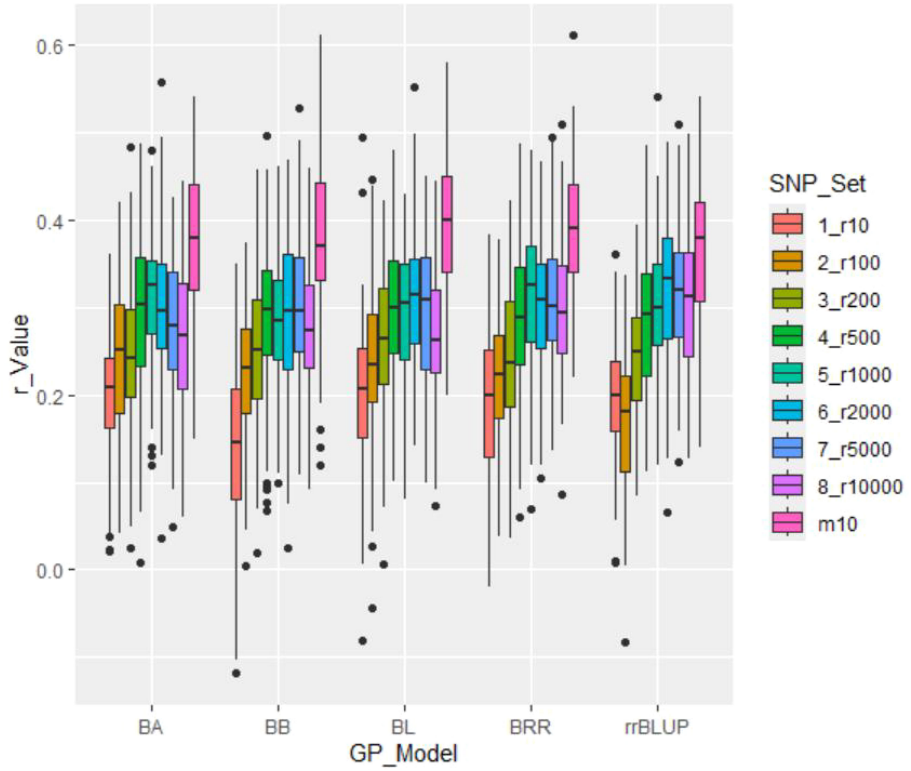


FIGURE 4
Genomic prediction (r-value) of salt tolerance using nine SNP sets: eight randomly selected SNP sets ranging from 10 (r10) to 10,000 SNPs (r10000), plus the GWAS-derived SNP marker sets (10 markers - m10). Predictions were estimated using five genomic prediction (GP) models: BA, BB, BL, BRR, and rrBLUP.

adverse environmental conditions, with salt stress being a significant abiotic factor that severely impacts soybean production and poses a substantial threat to agricultural productivity (Leng et al., 2021; van Zelm et al., 2020). Identifying genes associated with salt stress tolerance is essential for developing salt-tolerant soybean varieties and improving soybean yields. Although some genes regulating salt tolerance traits have been reported, research on soybean salt tolerance remains insufficiently

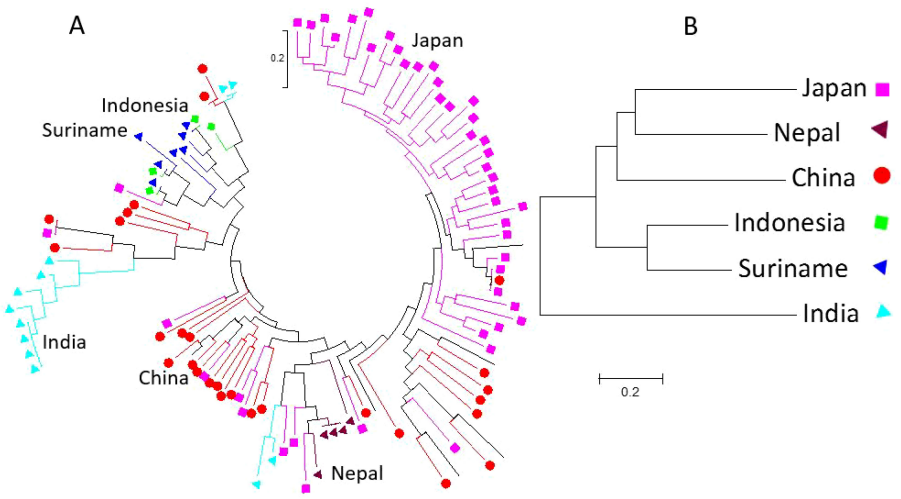


FIGURE 5
The non-taxon ring phylogenetic tree of 104 salt-tolerant soybean germplasms was constructed using the Maximum Likelihood (ML) method in MEGA 7, based on 6000 randomly selected SNPs distributed across 20 soybean genome chromosomes **(A)**. The colored shapes and branches represent germplasms from one of the six countries: Japan, Nepal, China, India, Suriname, and Indonesia. The traditional phylogenetic tree of soybean germplasm from the six countries is shown in **(B)**.

comprehensive. Therefore, this study aims to analyze the salt tolerance of 563 soybean germplasms resource from the USDA GRIN, identify salt tolerance-related genes, and conduct genomic predictions. These efforts are vital for selecting salt-tolerant soybean germplasm and breeding salt-tolerant soybean varieties.

In recent years, numerous QTLs associated with soybean salt tolerance have been identified. However, due to the complex nature of soybean salt tolerance, which is controlled by multiple genes, the related loci and candidate genes identified vary across different populations or using different analytical methods. This study conducted a GWAS on 34,181 SNP markers, identifying 10 SNPs associated with salt tolerance, located on chromosome 1, 2, 3, 7, and 16. In both GAPIT models (GLM and FarmCPU), a locus on chromosome 3 showed LOD scores exceeding 5.83 and t-test values greater than 8.9, indicating a robust QTL for salt tolerance. Four genes—*Glyma.03G230400*, *Glyma.03G230500*, *Glyma.03G230600*, and *Glyma.03G230700*—are closely linked to SNP markers Gm03_43213208_ss715586397 and Gm03_43220331_ss715586399 within 4 kb. Previous studies have also identified salt-tolerant genes in this region (Kan et al., 2015; Patil et al., 2016; Zeng et al., 2017). On chromosome 2, three SNP markers—Gm02_36878905_ss715582154, Gm02_36940321_ss715582156, and Gm02_36991983_ss715582157—within the 36,878,905 bp to 36,991,983 bp region exhibited LOD scores greater than 6.7 and t-test values exceeding 4.1, suggesting a salt tolerance QTL. The gene *Glyma.02G195400* is closely linked to SNP marker Gm02_36878905_ss715582154, within 4 kb. Similar QTLs have been reported by Zeng et al. (2017) and Kan et al. (2015). On chromosome 1, three SNP markers—Gm01_4306329_ss715579436, Gm01_4312808_ss715579441, and Gm01_4336306_ss715579451—located within the 4306329 bp to 4336306 bp region showed LOD scores greater than 6.5 and significant t-test values, indicating a salt tolerance QTL. Four genes—*Glyma.01G039600*, *Glyma.01G039700*, *Glyma.01G039800*, and *Glyma.01G040000*—are closely linked to these markers within 5 kb. Similar QTLs have been identified in rice by Pandit et al. (2010). On chromosome 7, SNP marker Gm07_41776639_ss715598058 at 41,776,639 bp showed an LOD score of 6.44 and a t-test value of 5.94, indicating significant association with salt tolerance. This region was also identified by Zeng et al. (2017). The SNP marker Gm16_7722217_ss715625494 on chromosome 16, located at 7,722,217 bp, exhibited an LOD score of 5.95 and a t-test value of 2.75, suggesting a salt tolerance QTL. Two genes—*Glyma.16G076200* and *Glyma.16G076300*—are closely linked to this marker within 3 kb. Similar QTLs in this region were identified by Zeng et al. (2017).

In this study, the accuracy of GP was evaluated by assessing the correlation coefficient (r) between the GEBV and the observed values. (Bhattarai et al., 2022; Joshi et al., 2022) Initially, salt tolerance of 563 soybean germplasms were predicted using three different genomic prediction models: maBLUP, gBLUP, and sBLUP for itself by cross-population prediction. The r values obtained from these models were 0.46, 0.60, and 0.44, respectively, indicating that genomic selection for salt tolerance is effective. Subsequently, cross-prediction was conducted using randomly selected SNP markers and GWAS-derived SNP marker sets. The results showed that the r-values were relatively higher for the GWAS-derived SNP marker sets. All five GP models—BA, BB, BL, BRR, and rrBLUP—exhibited similar r values, demonstrating that each model is effective for selecting salt tolerance through GS. These findings suggested that GP and salt tolerance selection can be effectively utilized in soybean breeding through MAS and GS.

Conclusion

Through GWAS analysis of 150 tolerant and 413 susceptible germplasms with 34,181 SNP loci, we identified 10 SNPs associated with salt tolerance: four SNP markers on chromosome 1 (Gm01_4306329_ss715579436, Gm01_4312808_ss715579441, and Gm01_4336306_ss715579451), three markers on chromosome 2 (Gm02_36878905_ss715582154, Gm02_36940321_ss715582156, and Gm02_36991983_ss715582157), two markers on chromosome 3 (Gm03_43213208_ss715586397 and Gm03_43220331_ss715586399), and one marker each on chromosomes 7 and 16 (Gm07_41776639_ss715598058 and Gm16_7722217_ss715625494, respectively). We assessed the accuracy of GP by examining the correlation coefficients (r) between GEBV and observed values. Using different GP models and SNP sets, we observed that r-values were up to 0.4 when using significant SNP markers derived from GWAS. The information provided valuable references for selecting and breeding soybean varieties with enhanced salt tolerance.

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

RX: Writing – original draft, Writing – review & editing. QY: Writing – original draft, Writing – review & editing. ZL: Writing – review & editing. XS: Writing – review & editing. XW: Writing – review & editing. YC: Writing – review & editing. XD: Writing – review & editing. QG: Writing – review & editing. DH: Writing – review & editing. AS: Writing – review & editing. PT: Writing – review & editing. LY: Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study is

supported by China Agriculture Research System of MOF and MARA (CARS-04-PS06), the National Key Research and Development Program (2023YFD2301505), Hebei Agriculture Research System (HBCT2023040101), Introduction and Screening of Soybean Germplasm Resources in Bashang, coupled with the Breeding of Novel Varieties (2023KJCXZX-LYS-23), Scientific and Technological Innovation Talent Team Construction (C24R0302; C24R0308).

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.2024.1494551/full#supplementary-material

References

Beecher, H. G. (1994). Effects of saline irrigation water on soybean yield and soil salinity in the Murrumbidgee Valley. *Aust. J. Exp. Agric.* 34, 85–91. doi: 10.1071/EA9940085

Bhattarai, G., Shi, A., Mou, B., and Correll, J. C. (2022). Resequencing worldwide spinach germplasm for identification of field resistance QTLs to downy mildew and assessment of genomic selection methods. *Horticulture Res.* 9, uhac205. doi: 10.1093/hr/uhac205

Chen, H., Cui, S., Fu, S., Gai, J., and Yu, D. (2008). Identification of quantitative trait loci associated with salt tolerance during seedling growth in soybean (*Glycine max L.*). *Crop Pasture Sci.* 59, 1086–1091. doi: 10.1071/AR08104

Du, H., Fang, C., Li, Y., Kong, F., and Liu, B. (2023). Understandings and future challenges in soybean functional genomics and molecular breeding. *J. Integr. Plant Biol.* 65, 468–495. doi: 10.1111/jipb.13433

Duhnen, A., Gras, A., Teyssedre, S., Romestant, M., Claustres, B., Daydé, J., et al. (2017). Genomic selection for yield and seed protein content in soybean: a study of breeding program data and assessment of prediction accuracy. *Crop Sci.* 57, 1325–1337. doi: 10.2135/cropsci2016.06.0496

Endelman, J. B. (2011). Ridge regression and other kernels for genomic selection with R package rrBLUP. *Plant Genome.* 4, 250–255. doi: 10.1104/pp.74.1.72

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

Hickey, J. M., Chiurugwi, T., Mackay, I., Powell, W., Implementing Genomic Selection in CGIAR Breeding Programs Workshop Participants (2017). Genomic prediction unifies animal and plant breeding programs to form platforms for biological discovery. *Nat. Genet.* 49, 1297–1303. doi: 10.1038/ng.3920

Huang, L. (2013). Genome-wide association mapping identifies QTLs and candidate genes for salt tolerance in soybean. [dissertation/master's thesis]. Fayetteville (AR): University of Arkansas.

Jarquín, D., Kocak, K., Posadas, L., Hyma, K., Jedlicka, J., Graef, G., et al. (2014). Genotyping by sequencing for genomic prediction in a soybean breeding population. *BMC Genomics* 15, 740. doi: 10.1186/1471-2164-15-740

Jarquín, D., Lemes da Silva, C., Gaynor, R. C., Poland, J., Fritz, A., Howard, R., et al. (2017). Increasing genomic-enabled prediction accuracy by modeling genotype × environment interactions in Kansas wheat. *Plant Genome.* 10. doi: 10.3835/plantgenome2016.12.0130

Joshi, V., Shi, A., Mishra, A. K., Gill, H., and DiPiazza, J. (2022). Genetic dissection of nitrogen induced changes in the shoot and root biomass of spinach. *Sci. Rep.* 12, 13751. doi: 10.1038/s41598-022-18134-7

Kan, G., Ning, L., Li, Y., Hu, Z., Zhang, W., He, X., et al. (2016).). Identification of novel loci for salt stress at the seed germination stage in soybean. *Breed. Sci.* 66, 530–541. doi: 10.1270/jsbbs.15147

Kan, G., Zhang, W., Yang, W., Ma, D., Zhang, D., Hao, D., et al. (2015). Association mapping of soybean seed germination under salt stress. *Mol. Gen. Genomics* 290, 2147–2162. doi: 10.1007/s00438-015-1066-y

Keller, B., Ariza-Suarez, D., de la Hoz, J., Aparicio, J. S., Portilla-Benavides, A. E., Buendia, H. F., et al. (2020). Genomic prediction of agronomic traits in common bean (*Phaseolus vulgaris* L.) under environmental stress. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.01001

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

Leng, Z. X., Liu, Y., Chen, Z. Y., Guo, J., Chen, J., and Zhou, Y. B. (2021). Genome-wide analysis of the duf4228 family in soybean and functional identification of *GmDUF4228-70* in response to drought and salt stresses. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.628299

Lu, L., Wei, W., Tao, J. J., Lu, X., Bian, X. H., Hu, Y., et al. (2021). Nuclear factor Y subunit GmNFYA competes with GmHDA13 for interaction with GmFVE to positively regulate salt tolerance in soybean. *Plant Biotechnol. J.* 19, 2362–2379. doi: 10.1111/pbi.13668

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, 428–436. doi: 10.1038/s41588-020-0604-7

Ondrasek, G., Rengel, Z., and Veres, S. (2011). "Soil salinisation and salt stress in crop production," in *Abiotic Stress in Plants: Mechanisms and Adaptations*. Eds. A. K. Shanker and B. Venkateswarlu (INTECH Press, Rijeka), 171–190. doi: 10.5772/22248

Pandit, A., Rai, V., Bal, S., Sinha, S., Kumar, V., Chauhan, M., et al. (2010). Combining QTL mapping and transcriptome profiling of bulked RILs for identification of functional polymorphism for salt tolerance genes in rice (Oryza sativa L.). *Mol. Genet. Genomics* 284, 121–136. doi: 10.1007/s00438-010-0551-6

Patil, G., Do, T., Vuong, T. D., Valliyodan, B., Lee, J. D., Chaudhary, J., et al. (2016). Genomic-assisted haplotype analysis and the development of high-throughput SNP markers for salinity tolerance in soybean. *Sci. Rep.* 6, 19199. doi: 10.1038/srep19199

Sedivy, E. J., Wu, F., and Hanzawa, Y. (2017). Soybean domestication: the origin, genetic architecture and molecular bases. *New Phytol.* 214, 539–553. doi: 10.1111/nph.14418

Shi, A., Bhattarai, G., Xiong, H., Avila, C. A., Feng, C., Liu, B., et al. (2022). Genome-wide association study and genomic prediction of white rust resistance in USDA GRIN spinach germplasm. *Hortic. Res.* 9, uhac069. doi: 10.1093/hr/uhac069

Shi, A., Buckley, B., Mou, B. Q., Motes, D., Morris, J. B., Ma, J. B., et al. (2016). Association analysis of cowpea bacterial blight resistance in USDA cowpea germplasm. *Euphytica* 208, 143–155. doi: 10.1007/s10681-015-1610-1

Shi, A., Gepts, P., Song, Q., Xiong, H., Michaels, T. E., and Chen, S. (2021). Genome-wide association study and genomic prediction for soybean cyst nematode resistance in USDA common bean (phaseolus vulgaris) core collection. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.624156

Shi, A., Qin, J., Mou, B. Q., Correll, J., Weng, Y. J., Brenner, D., et al. (2017). Genetic diversity and population structure analysis of spinach by singlenucleotide polymorphisms identified through genotyping-by-sequencing. *PloS One* 12, e0188745. doi: 10.1371/journal.pone.0188745

Shikha, M., Kanika, A., Rao, A. R., Mallikarjuna, M. G., Gupta, H. S., and Nepolean, T. (2017). Genomic selection for drought tolerance using genome-wide SNPs in maize. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.00550

Singleton, P. W., and Bohlool, B. B. (1984). Effect of salinity on nodule formation by soybean. *Plant Physiol.* 74, 72–76. doi: 10.1104/pp.74.1.72

Song, Q., Hyten, D. L., Jia, G., Quigley, C. V., Fickus, E. W., Nelson, R. L., et al. (2013). Development and evaluation of SoySNP50K, a high-density genotyping array for soybean. *PloS One* 8, e54985. doi: 10.1371/journal.pone.0054985

Song, Q., Hyten, D. L., Jia, G., Quigley, C. V., Fickus, E. W., Nelson, R. L., et al. (2015). Fingerprinting soybean germplasm and its utility in genomic research. *G3 (Bethesda)* 5, 1999–2006. doi: 10.1534/g3.115.019000

van Zelm, E., Zhang, Y., and Testerink, C. (2020). Salt tolerance mechanisms of plants. *Annu. Rev. Plant Biol.* 71, 403–433. doi: 10.1146/annurev-arplant-050718-100005

Wang, D. G., Yang, Y., Hu, G. Y., and Huang, Z. P. (2023). Comparative analysis of yield traits of soybean lines (species) in southern Huang-Huai region from 2016 to 2021. *Chin. J. Oil-Bearing Crops* 1-9. doi: 10.19802/j.iSSN.1007-9084.2022256

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

Zeng, A., Chen, P., Chen, P., Korth, K. L., Hancock, F., Pereira, A., et al. (2017). Genome-wide association study (GWAS) of salt tolerance in worldwide soybean germplasm lines. *Mol. Breed.* 37, 1–14. doi: 10.1007/s11032-017-0634-8

Zhang, H. Y., Li, C. Y., Davis, E. L., Wang, J. S., Griffin, J. D., Kofsky, J., et al. (2016). Genome-wide association study of resistance to soybean cyst nematode (Heterodera glycines) HG Type 2.5.7 in wild soybean (*Glycine soja*). *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.01214

Zhang, J. P., Song, Q. J., Cregan, P. B., and Jiang, G. L. (2016). Genomewide association study, genomic prediction and marker-assisted selection for seed weight in soybean (*Glycine max*). *Theor. Appl. Genet.* 129, 117–130. doi: 10.1007/s00122-015-2614-x

Zhang, W., Liao, X., Cui, Y., Ma, W., Zhang, X., Du, H., et al. (2019). A cation diffusion facilitator, *GmCDF1*, negatively regulates salt tolerance in soybean. *PloS Genet.* 15, e1007798. doi: 10.1371/journal.pgen.1007798

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 27 January 2025
DOI 10.3389/fpls.2024.1518829

OPEN ACCESS

EDITED BY
Huatao Chen,
Jiangsu Academy of Agricultural Sciences (JAAS), China

REVIEWED BY
Yingpeng Han,
Northeast Agricultural University, China
Wenlong Li,
Hebei Agricultural University, China

*CORRESPONDENCE
Kai Li
kail@njau.edu.cn
Junyi Gai
sri@njau.edu.cn

RECEIVED 29 October 2024
ACCEPTED 27 November 2024
PUBLISHED 27 January 2025

CITATION
Raza MM, Jia H, Razzaq MK, Li B, Li K and Gai J (2025) Identification and functional validation of a new gene conferring resistance to *Soybean Mosaic Virus* strains SC4 and SC20 in soybean. *Front. Plant Sci.* 15:1518829. doi: 10.3389/fpls.2024.1518829

Identification and functional validation of a new gene conferring resistance to *Soybean Mosaic Virus* strains SC4 and SC20 in soybean

Muhammad Muzzafar Raza, Huiying Jia, Muhammad Khuram Razzaq, Bowen Li, Kai Li* and Junyi Gai*

Soybean Research Institute & MARA National Center for Soybean Improvement & Ministry of Agriculture and Rural Affairs (MARA) Key Laboratory of Biology and Genetic Improvement of Soybean (General) & State Innovation Platform for Integrated Production and Education in Soybean Bio-breeding & State Key Laboratory for Crop Genetics and Germplasm Enhancement & Jiangsu Collaborative Innovation Center for Modern Crop Production, Nanjing Agricultural University, Nanjing, Jiangsu, China

Soybean Mosaic Virus (SMV) poses a serious threat to soybean production, often resulting in considerable yield losses or complete crop failure, particularly if infection occurs during early growth stages. While several SMV resistance genes have been identified, the genetic basis of resistance to certain strains remains poorly understood. Among the 22 SMV strains, SC4 and SC20 are considered pathogenic in Central China. Dominant genes resistant to SC4 (*Rsc4*) on Chr.14 in Dabaima and to SC20 (*Rsc20*) on Chr.13 in Qihuang-1 have been identified. Kefeng-1 is resistant to SC4 and SC20. This study aimed to determine whether the resistance to SC4 and SC20 in Kefeng-1 was identical and whether *Rsc4* and *Rsc20* in Dabaima and Qihuang-1 are also present in Kefeng-1 due to translocation. Mendelian experiments using F_1, F_2, and recombinant inbred lines ($RIL_{3:8}$) of Kefeng-1 (resistant) and NN1138-2 (susceptible) indicated a single dominant gene inheritance pattern in SC4 and SC20, respectively. Linkage mapping showed two loci for SC4 and SC20 in neighboring single nucleotide polymorphism linkage disequilibrium blocks (SNPLDB) marker regions of 253 kb and 375 kb, respectively, in Kefeng-1. Association between SNPs in possible gene regions of Kefeng-1 and resistance data showed SNP11692903 jointly as the most significant SNP, exhibiting the highest χ^2 value. By comparing SNP11692903 to possible gene sequences in the coding region, *Glyma02g13380* was identified as a joint candidate gene. The results were validated using qRT-PCR, virus induced gene silencing (VIGS), and gene-sequence. Therefore, the two Mendelian genes on chromosome 2 in Kefeng-1 responsible for SC4 and SC20 resistance are unique genes, different from *Rsc4* in Dabaima and *Rsc20* in Qihuang-1. Hence, one gene is involved in resistance

toward two SMV strains resistance. This result challenged our previous hypothesis of a single dominant gene responsible for resistance against a single strain and underscored the potential for using multiple resistance sources aimed at enhancing SMV resistance in breeding practices.

KEYWORDS

Mendelian inheritance, Soybean Mosaic Virus, single nucleotide polymorphism, resistance, linkage mapping

1 Introduction

Soybean is an economically vital grain legume because of its protein and edible oil of excellent quality that are used in industry and routine human affairs. One of the major risks for soybean production in China and worldwide is *Soybean Mosaic Virus* (SMV). Viral infection during the early growth stages of crops may lead to maximum yield loss or complete crop failure (Hill, 2003). If infection occurs within three weeks of sowing, it can cause up to 100% crop failure (Hill and Whitham, 2014).

Several SMV strains have been reported worldwide; however, the pathogenic relationship among these strains has not yet been completely elucidated. In Japan, five strains, represented by A, B, C, D, and E, have been reported (Takahashi et al., 1980). In the USA, seven strains (G1, G2, G3, G4, G5, G6, and G7) were identified by Cho and Goodman (1982), and this grouping was modified by Chen and Choi (2007). Three discrete resistance loci for these strains on Chromosome 2, 13 and 14 from their natural sources have been discovered (Jeong and Saghai Maroof, 2004). In China, 22 SMV strains (SC1-SC22) have been identified in ten varieties of/differential hosts (Li et al., 2010; Wang et al., 2013; Zhi et al., 2005). Host plant resistance (HPR) is a cost-effective and eco-friendly technique for combating SMV. Numerous sources of SMV resistance have been identified in soybeans, most of which are resistant to a few strains. In most cases, the resistance is conferred by a single dominant gene. Until now, many self-governing resistance loci with diverse SMV strain specifications have been recognized. To date, three autonomous loci; *Rsv1*, *Rsv*3, and *Rsv*4 in the USA and several R_{SC} loci in China have been reported and identified that confer SMV resistance (Liu et al., 2016). However, only a few resistance genes have been cloned, and their characteristics are yet to be identified.

A lot of research on SMV resistance has been conducted worldwide such as in cultivar "Heinong 84" a single dominant gene on Chr.13 which is allele of *Rsv1* is controlling resistance against strain N3 (Li et al., 2022). Similarly, another SMV resistant novel locus R_{smv}*-11* for SC7 was identified (Zhang et al., 2022). *Rsv1*, *Rsv3* loci were discovered in Egyptian soybean cultivars, SMV SC3 was studied in NN1138-2 and Parasite Driven Resistance (PDR) for SMV was studied (Eid et al., 2023; Liu et al., 2023; Yang et al., 2024).

One of the basic requirements for the development of SMV-resistant soybean cultivars using molecular breeding techniques is the identification of genomic regions with the most probable candidate disease-resistant genes. Linkage mapping is an excellent technique for gene identification, especially when combined with map-based cloning (Fu et al., 2006; Hayes et al., 2004; Hwang et al., 2006; Li et al., 2006), but the accuracy of this technique depends on map resolution. Recently, single-nucleotide polymorphism (SNP) markers have become popular because they focus on molecular breeding and are suitable for high-throughput recognition setups and platforms (Ganal et al., 2009). SNP markers are commonly used to develop high-quality resolution maps. These SNP markers are common sources of variability in the genome and are used to measure genetic diversity and population structure. Different public databases comprise extensive information on SNPs in numerous plant species, including rice, wheat, maize, and soybean (Akond et al., 2013; Cavanagh et al., 2013; Ganal et al., 2011; Jones et al., 2009; McNally et al., 2009; Rostoks et al., 2005; Song et al., 2013; Tian et al., 2015).

Among the 22 SMV strains, SC4 and SC20 are moderately pathogenic strains prevalent in the Huang-Huai and Yangtze River Valleys in China (Li et al., 2010). R_{SC4} locus has been previously reported in Dabaima and contains three candidate resistance genes on chromosome (Chr.) 14 (Wang et al., 2011). In contrast, R_{SC20} has been reported on Chr.13 in Qihuang-1 (Karthikeyan et al., 2018) as a single dominant resistance gene that has not yet been cloned. Based on these and previous studies, we recognized that a single dominant gene is responsible for single-strain resistance. Recently, we found that Kefeng-1 also shows resistance against SC4 and SC20; however, the responsible elements are not located on Chr.14 and Chr.13 (Wang et al., 2011). Therefore, this study aimed to investigate whether the resistance to SC4 and SC20 in Kefeng-1 are conferred by distinct genes, determine their location, and clarify whether the newly found genes resistant to SC4 and SC20 were distinct or identical to R_{SC4} and R_{SC20} in Dabaima and Qihuang-1 due to translocation. The objectives of this study were to identify resistance genes for both strains in Kefeng-1 using the techniques of Mendelian inheritance, gene mapping, gene fine mapping, quantitative reverse transcription polymerase chain reaction (qRT-PCR), virus-induced gene silencing (VIGS), and gene sequencing to gain a thorough understanding and improve our resistance breeding strategy (Zhan et al., 2006).

2 Materials and methods

2.1 Plant genetic materials

The soybean cultivar NN1138-2 was susceptible (S), and Kefeng-1 was resistant to the SMV strains SC4 and SC20. To investigate the inheritance patterns of resistance, F_1, F_2, and $RIL_{3:8}$ (comprising 427 RILs) derived from a cross between Kefeng-1 and NN1138-2 were used. For fine mapping of candidate resistance genes, only the RIL population was used. The parental lines were crossed to acquire F_1 at the Dangtu Research Station, National Center for Soybean Improvement (NCSI), Nanjing Agricultural University, China. The F_1 plants were self-fertilized to obtain F_2 plants in Hainan province. To derive the RIL population, a single-seed descent procedure was used to obtain the $F_{3:8}$ generation. Both parents and their derived populations were cultivated for screening in plastic pots (w20×h20 cm) in a fully controlled, conditioned glass house at the Baima Agriculture Station, Nanjing Agricultural University, China.

2.2 Preparation and maintenance of SMV inoculum

The SMV strains SC4 and SC20 were supplied by NCSI. The SMV isolates were maintained separately on NN1138-2. Plants at the first true-leaf stage were successively inoculated under controlled conditions in an aphid-free glass house. Inoculation performed mechanically as described earlier (Che et al., 2020). Infected fresh leaves of NN1138-2 were ground using a pestle in a mortar containing 0.01 mol/L of sodium phosphate buffer (4–5 ml/g of leaf tissue; 7.2 pH) to prepare the inoculum. The mortar containing the inoculum was placed on ice during inoculation. Before inoculation, the leaves were dusted with 600 mesh carborundum powder as an abrasive and rinsed with tap water after inoculation. The SC4 and SC20 inocula were verified by the real inoculation of a set of ten differential soybean hosts, as described by Li et al. (2010) (Supplementary Table 1).

2.3 Inoculum application and disease assay

The plants of F_1, F_2, and RIL populations with their parental lines (ten plants per line) were mechanically inoculated separately with SC4 and SC20 inoculums at the true-leaf unfolding stage (seven-to-ten days after sowing). The pathological response was recorded based on optical observance with a time interval of 10 days until 40 days after inoculation as described by Karthikeyan et al. (2017). Mosaic and necrosis-like symptoms were observed on newly emerging leaves after inoculation. Categorization criteria for susceptibility and resistance to specific SMV strains were established based on symptomatic and symptomless plants. Plants that did not develop any salient evidence of infection after inoculation and looked like uninoculated plants and those with flimsy necrotic stripes, minor necrotic lesions, and fainted mosaic evidence with less than 10% rate of incidence were characterized in the resistance category, whereas plants that displayed symptoms such as mosaic with or without necrotic leaves and general necrosis with an incidence rate greater than 10% were regarded as susceptible.

2.4 Statistical analysis

The chi-square test was used to test the phenotypic assessment of segregation in the F_2 and RIL populations for goodness of fit to the Mendelian ratio. The values of Chi-square and p were obtained using the SAS software (Version 9.4).

2.5 Genotypic resequencing and gene mapping

Genotypic data of the RIL population (Kefeng-1×NN1138-2) comprising 3683 SNPLDB markers were used to track the resistant genomic regions provided by NCSI. The construction procedure for the SNPLDB markers has been outlined by He et al. (2017). The protocol is described in brief as follow: Genomic DNA from both parents and the RIL population was isolated from young leaves following the cetyltrimethylammonium bromide (CTAB) methodology (Murray and Thompson, 1980). All sequence libraries were constructed using the Illumina HiSeq 2000 protocol. The SNP calling and genotyping were performed using component combining Burrows-Wheeler Aligner (Li and Durbin, 2009) and Sequence Alignment/Map tools (SAMtools) (Li et al., 2009). SNPs with deletions of less than or equal to 20%, error rates of less than or equal to 1%, and heterozygosity rates of less than 5% were deleted, and NPUTE was used to fill in the SNPs (Roberts et al., 2007). An SNPLDB marker was structured with multiple haplotypes as alleles by grouping the SNPs within an LD block. Each genotype was resolved using related haplotypes at each locus. The LD values between multi-allelic SNPLDBs were calculated as previously described by Farnir et al. (2000).

Linkage map construction was performed using JoinMap 4.1. A total of 3683 SNPLDB markers distributed among all 20 chromosomes of soybean were used to construct a linkage map. The segregation of markers on each chromosome was 243 on Chr.1, 157 on Chr.2, 198 on Chr.3, 280 on Chr.4, 126 on Chr.5, 112 on Chr.6, 114, 225, 170, 267, 134, 75, 157, 216, 167, 206, 197, 322, 167 and 13 on Chr. 7, 8. 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 and 20 respectively. A logarithm of Odds (LOD) threshold of 3.0 was used, and the marker distance was determined according to the Kosambi function (Kosambi, 2016).

2.6 Selection of candidate gene and association analysis

After genomic region identification, the total number of genes located in that region were selected from the Williams 82 reference

genome (GlymaWm82.a1.v1), and homologous genes in *Arabidopsis thaliana* with their functional annotation were identified from soybase (https://www.soybase.org) and phytozome (https://www.phytozome.jgi.doe.gov) database. The whole-genome resequencing (WGRS) data from 427 families were used to accumulate the total SNPs in the identified genomic region. Genes possessing SNPs in their coding regions with nonsynonymous translations were selected. Finally, WGRS gene marker data were used to find the most plausible candidate gene, and the affiliation between gene markers and the phenotypic data set of all 427 RILs was inspected by the Chi-square test using software *R 3.2.3* at $p < 0.01$. We used the most probable SNP in the coding region to identify the most probable gene in the shortened region.

2.7 SC4 and SC20 resistance locus identification using WGRS gene-marker data

To obtain the most appropriate SC4 resistance gene, WGRS gene marker data from 427 families were used to construct a linkage map to identify SC4 resistance loci among the gene marker loci. Similarly, another linkage map was constructed for SC20.

2.8 qRT-PCR for expression analysis

Quantitative RT-PCR was performed to determine the candidate gene expression. TaqMan software (http://www.genscript.com) was used to design gene-specific primers from the respective coding sequences acquired from the SoyBase database (http://www.soybase.org/) (Supplementary Table 2). Tubulin was used as the internal reference gene. The expression differential of genes was quantified by the $2^{-\Delta\Delta\ Ct}$ method (Livak and Schmittgen, 2001). Each biological replicate contained three samples with three technical replicates.

2.9 VIGS vectors construction

After identifying candidate resistance genes, their functions were validated using VIGS. Bean Pod Mottle Virus (BPMV) BPMV-silencing vectors were constructed and inoculated as previously described (Zhang et al., 2009). The restriction enzymes *BamH*I and *Sal*I (NEB, Beijing, China) were mixed with pBPMV-IA-V2-R2, and a 300 bp region of all candidate genes (*Glyma02g13230, Glyma02g13380, Glyma02g13401, Glyma02g13470, Glyma02g13570*, and *Glyma02g13630*) was amplified from Kefeng-1. The BPMV-linearized vectors and gene fragments were combined using recombination homology. The recombinant plasmids and pBPMV-IA-R1M were separately multiplied and mixed in equal ratios to mechanically inoculate the NN1138-2 plants.

2.10 Silencing efficiency estimation and SC4 and SC20 accumulation detection by qRT-PCR and ELISA

NN1138-2 leaves showing SMV symptoms after inoculation with silencing vectors were used for RNA extraction. Total RNA extraction and cDNA synthesis were performed according to the manufacturer's protocols (Accurate Biotechnology Co., Ltd., Changsha, China; Vazyme Biotech, Nanjing, China). RT-PCR was performed using specific primers to confirm the insertion of the gene fragments. After confirmation of the inserted gene fragment, diseased leaves with silencing vectors for the candidate genes were inoculated on the first true leaves of Kefeng-1, and SMV strains SC4 and SC20 were inoculated separately on the first trifoliate leaves. Ten days after SMV inoculation, upper uninoculated leaves were collected for the detection of the silencing efficiency of the concerned vectors and SMV accumulation by qRT-PCR and double-antibody sandwich enzyme-linked immunosorbent assay (DAS-ELISA). The expression differential of respective genes was quantified using the $2^{-\Delta\Delta Ct}$ method. Each biological replicate comprised three samples with three technical replicates. The primers used for gene fragment insertion confirmation, silencing efficiency, and CP content are listed in Supplementary Table 3.

3 Results

3.1 Mendelian inheritance mechanisms of resistance toward SC4 and SC20 in Kefeng-1

Both parents, Kefeng-1 and NN1138-2, and their offspring were inoculated with SMV strains SC4 and SC20, which produced different symptoms. Kefeng-1 did not show any symptoms in response to either SMV strain, and the plants were similar to the non-inoculated plants in the control group. In contrast, plants of the second parent NN1138-2 showed stunted growth symptoms, reduced leaf size, and mosaic patterns on their leaves within 15–20 days post-inoculation (dpi) with SMV strains. Consequently, Kefeng-1 and F_1 generations were resistant to SC4 and SC20, whereas NN1138-2 was susceptible.

Polymorphisms were observed in both parents (Kefeng-1 and NN1138-2) after inoculation with SMV strains SC4 and SC20. All plants of the sensitive parent NN1138-2 were scrawny and scrubby in growth, with mosaic symptoms on their leaves, whereas immune parent plants remained symptomless, similar to uninoculated control plants, until 30 days of inoculation. F_1s of Kefeng-1 and NN1138-2 showed excellent immunity to SC4 and SC20, as evident from the dominant resistance pattern recorded in Kefeng-1. In total, 179 F_2 plants and 427 RILs obtained from parental crosses were inoculated with SC4. The phenotypic segregation ratio was 133R:46S (fitting 3:1 ratio) in F_2 whereas the phenotypic and genotypic ratios were 211R:216S (fitting1:1 ratio) in RILs, indicating the involvement of a single dominant gene in SC4

10.3389/fpls.2024.1518829

TABLE 1 Mendelian inheritance of SC4 and SC20 resistance in parents, F_1, F_2, and RILs .

Parent/ progeny	SC4						SC20					
	No. of plants/ lines		Total	Expected ratios	χ^2	*p*-Value	No. of plants/ lines		Total	Expected ratios	χ^2	*p*-Value
	R	S					R	S				
Kefeng-1	20	–	20				20	–	20			
NN1138-2	–	20	20				–	20	20			
F_1	25	–	25				10	–	10			
F_2	133	46	179	3:1	0.0466	0.8292	133	46	179	3:1	0.0466	0.8292
RILs	211	216	427	1:1	0.0585	0.8088	204	207	411	1:1	0.0219	0.8824

R for resistant, and S for susceptible.

resistance (Table 1). Similarly, 179 F_2 plants and 411 RILs were inoculated with SC20, which showed a phenotypic ratio of 133R:46S (fitting 3:1 ratio) in F_2 plants and 204R:207S (1:1 ratio) in RILs, with another dominant gene responsible for resistance to SC20 in Kefeng-1 (Table 1).

3.2 Mapping the genes conferring resistance to SC4 and SC20 using linkage analysis with SNPLDB as markers

We investigated the SC4 and SC20 resistance-related genomic regions, with an RIL population of 427 families. Two separate linkage maps covering all twenty chromosomes of soybean were constructed based on the phenotypic and genotypic data of 427 individual lines with 3683 SNPLDB markers using JoinMap 4.1 at a threshold LOD 3.0. SC4 resistance region was located at 208.2 cM between two SNPLDB markers from SNPLDB320-SNPLDB321 with a genetic position of 209.0–207.6 cM and SC20 was at 214.0 cM between SNPLDB319-SNPLDB320 positioned at 220.6–208.2cM on Chr.2. These SNPLDB markers showed a strong association with adjacent loci conferring resistance to SC4 and SC20, located at the physical positions 11693196–1944513 bp and 11486875–1693196bp with approximate genomic intervals of 253 kb and 375 kb, respectively (Figure 1). In accordance with the Williams 82 soybean reference genome GlymaWm82 model 1.1, 17 genes in the SC4 region and 25 genes in the SC20 region with some common genes were situated. Their ornamentation on Chr.2 is

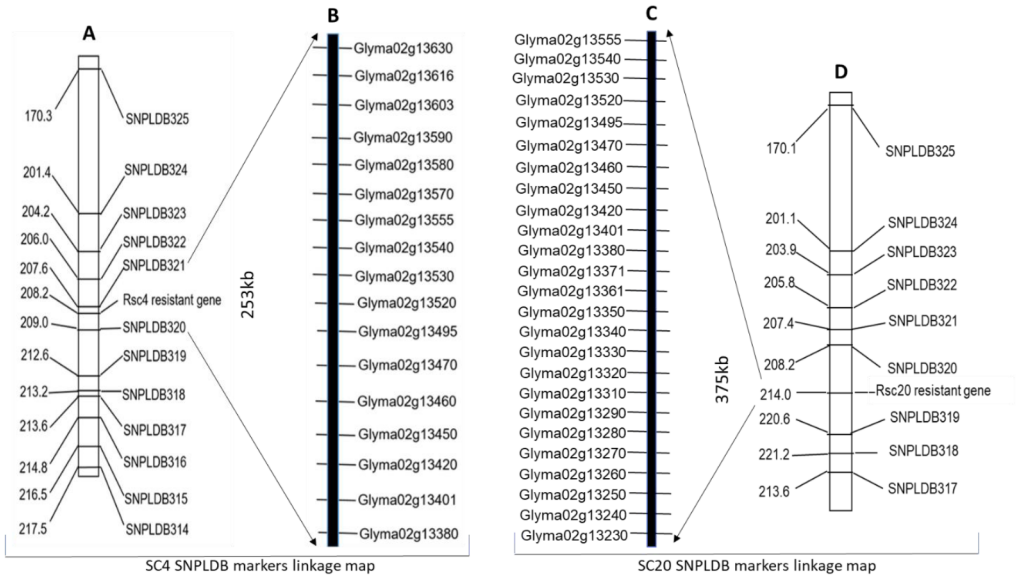


FIGURE 1
(A, D) SNPLDB linkage map showing SC4 and SC20 resistant loci between SNPLDB320-SNPLDB321 and SNPLDB319-SNPLDB320 on chromosome 2 constructed by WGRS genotypic and phenotypic data of 427 recombinant inbred lines, indicating genetic distance left side in centi Morgan (cM) and right-side names of SNPLDB markers, **(B, C)** genes found in 253-kb and 375-kb identified genomic regions retrieved from soybean William-82 reference genome GlymaWm82.a1. v1 in SoyBase (http://www.soybase.org).

10.3389/fpls.2024.1518829

TABLE 2 Eight SC4 and SC20 resistance nominee genes with their genomic data referred to William82 reference genome (GlymaWm82.a1. v1, http://www.soybase.org).

Sr. No	Gene	Gene Position	No. of SNPs	Allele	SNP Position
1	*Glyma02g13230*	11484118-11491037	1	G/A	11486875
2	*Glyma02g13361*	11647493-11663260	1	C/T	11662819
3	*Glyma02g13371*	11670695-11677275	1	T/A	11670897
4	*Glyma02g13380*	11692903-11694668	2	T/A	11693196
				A/G	11693900
5	*Glyma02g13401*	11729655-11746942	1	C/G	11739943
6	*Glyma02g13470*	11779259-11782486	1	T/A	11779670
7	*Glyma02g13570*	11890308-11895815	2	T/G	11893075
				T/G	11893936
8	*Glyma02g13630*	11943213-11849714	3	G/A	11944513
				C/T	11944870
				C/T	11946524

shown in Figure 1, and the functional annotation of *Arabidopsis* homologous genes is presented in Supplementary Table 4.

3.3 Candidate gene identification for resistance to SC4 and SC20 in the mapped regions through SNP association

We screened the most likely candidate genes for SC4 and SC20 resistance from all the genes in both regions over four gradual steps. In the first step, the WGRS was performed with the SNP dataset of all RIL families. A total of 26 SNPs in total were remunerated in the preoccupied SC4 and SC20 regions. Only 8 genes comprised 12 out of 26 SNPs, whereas 23 leftover genes did not comprise any SNP. *Glyma02g13230*, *Glyma02g13361*, *Glyma02g13371*, *Glyma02g13401*, and *Glyma02g13470* were each associated with one SNP; *Glyma02g13380* and *Glyma02g13570* were linked with two SNPs, whereas *Glyma02g13630* possessed three SNPs (Table 2).

In step two, we examined the coding and non-coding regions of eight SNP-holder genes. For SNP posture on genes, the genomic sequences of the concerned genes were retrieved from the SoyKB database (Soybean Knowledge Base, http://soykb.org). We found that SNPs in *Glyma02g13361*, *Glyma02g13371*, and *Glyma02g13401* were located in an intron and those in *Glyma02g13470* in a 5′UTR region. One of the two SNPs in *Glyma02g13570* was found in introns and the other in exons. Out of the three SNPs in *Glyma02g13630*, two were located in introns and one in exons, whereas *Glyma02g13230* and *Glyma02g13380* associated with one and two SNPs, respectively, had exons with nonsynonymous features of translation (Table 3).

Association analysis between gene markers from the WGRS SNPs data and phenotypic data of RILs for SC4 and SC20 resistance was performed in the third step. The results showed that *Glyma02g13630* was significantly correlated with SC4 resistance, *Glyma02g13230* with SC20 resistance, and *Glyma02g13380* and *Glyma02g13470* were significantly associated with both strains

TABLE 3 Four candidate genes indicating SNP location in their coding region and one gene in 5′UTR verified from SoyKB (http://soykb.org)..

Gene	No. of SNPs	SNP localized Zone	SNP associated CDS region	SNP position
Glyma02g13230	1	Exon	11485651-11488802	11486875
Glyma02g13380	2	Exon	11693175-11694242	11693196
		Exon		11693900
Glyma02g13470		5′UTR	5′UTR	11779670
Glyma02g13570	2	Exon	11893401-11894308	11893936
		Intron		11893075
Glyma02g13630	3	Exon	11944415-11944581	11944513
		Intron		11944870
		Intron		11946524

TABLE 4 Candidate genes associated with SC4 and SC20 resistance.

Sr. No.	Gene	*D.F*	χ^2	*P*-Value	χ^2	*P*-Value
			SC4		SC20	
1	*Glyma02g13230*	1	–	–	202.18	2.2e-16
2	*Glyma02g13380*	1	325.71	2.2e-16	297.8	2.2e-16
3	*Glyma02g13470*	1	311.93	2.2e-16	285.42	2.2e-16
4	*Glyma02g13630*	1	305.42	2.2e-16	–	–

(Table 4). Finally, in the fourth step, two separate linkage maps covering all twenty chromosomes of soybean were constructed using phenotypic data of 427 individual lines for SC4 and SC20 resistance and WGRS SNP gene markers data with the help of JoinMap 4.1 at a threshold level of LOD 3. Both SC4 and SC20 resistance loci were located between SNP11692903 and SNP14449821 on Chr.2 at 603.3 cM and 272.3 cM, respectively. SNP11692903, within the gene marker *Glyma02g13380*, was the closest to the SC4 and SC20 resistance loci (Figure 2). Eventually, based on polymorphisms (2 SNPs) in the coding region, a highly significant correlation with the resistance trait, and location nearest to the resistance loci, *Glyma02g13380* was considered to be the best candidate gene for SC4 and SC20 resistance in Kefeng-1. Furthermore, we hypothesized that the two different single dominant locus controlling SC4 and SC20 resistance, depicted in Mendelian inheritance, was actually the same gene that controlled the resistance of both strains.

3.4 Validation of *Glyma02g13380* through expression profile analysis

All genes with polymorphisms (five from the SC4 resistance region in Figure 3, and six from SC20 with three common genes in Figure 4) were assessed for their differential expression. The expression profiles of the genes on the groundwork of SNP enrichment were formulated with the help of quantitative real-time PCR using SC4 and SC20 separately inoculated leaves of both parental genotypes (Kefeng-1 and NN1138-2).

All five genes from the SC4 resistance region behaved differently in the expression analysis. *Glyma02g13380* seemed to exhibit a higher level of upregulated expression between the range of 2.84- to 24.78-folds in the resistant genotype at most time frames, especially at 2, 4, 8, 12, and 24 hours-post inoculation (hpi), whereas it was downregulated in susceptible parents at the same time points, except for 24 hpi (4.81 folds), when it was upregulated. Highly increased expression in the resistant parent was observed at 8 hpi with 24.78 folds, which is the statistically significant highest expression, followed by 19.96 folds at 4 hpi, 10.77 folds at 2 hpi, 9.94 folds at 12 hpi, and 9.07 folds at 24 hpi (Figure 3). *Glyma02g13401* and *Glyma02g13470* were expressed equally in both parents at most time points, excluding 4, 8, and 12 hpi in *Glyma02g13401* and 72 hpi in *Glyma02g13470*, where both genes were upregulated in the susceptible parental genotype. In *Glyma02g13570* up-regulation was recorded at 2, 4, and 8 h post-inoculation in the susceptible genotype Nannong 1138-2. In contrast, at 24 hpi, a higher upregulation was observed in the resistant parent Kefeng-1 compared to the opposite parent. This gene was downregulated at 2 hpi in Kefeng-1 and 48 hpi in both parental lines. Similarly, *Glyma02g13630* is

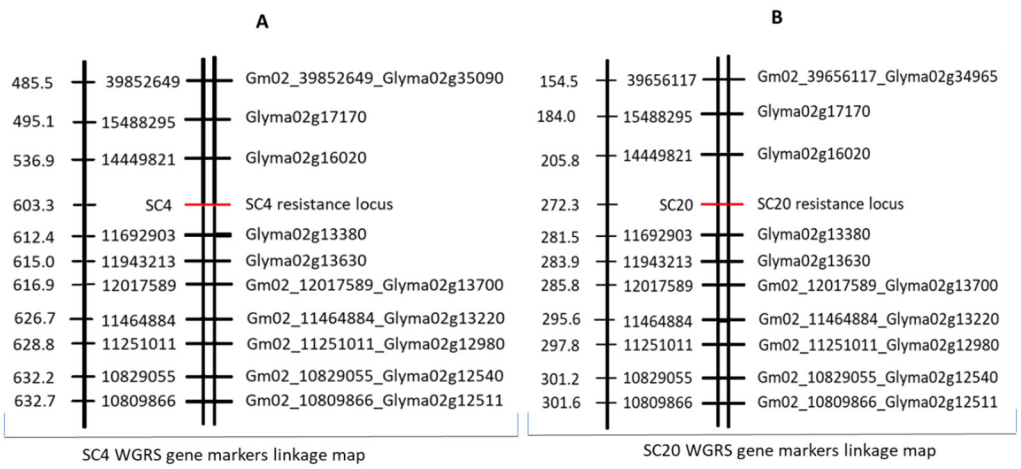


FIGURE 2
(A) linkage map for SC4 and **(B)** for SC20 showing resistant loci between two gene markers loci on chromosome 2 constructed by WGRS gene markers data and phenotypic data for SC4 and SC20 resistance of 427 recombinant inbred lines (RILs) indicating genetic distance in centi Morgan (cM) on its left side, SNP position in its middle, and gene markers on right-side.

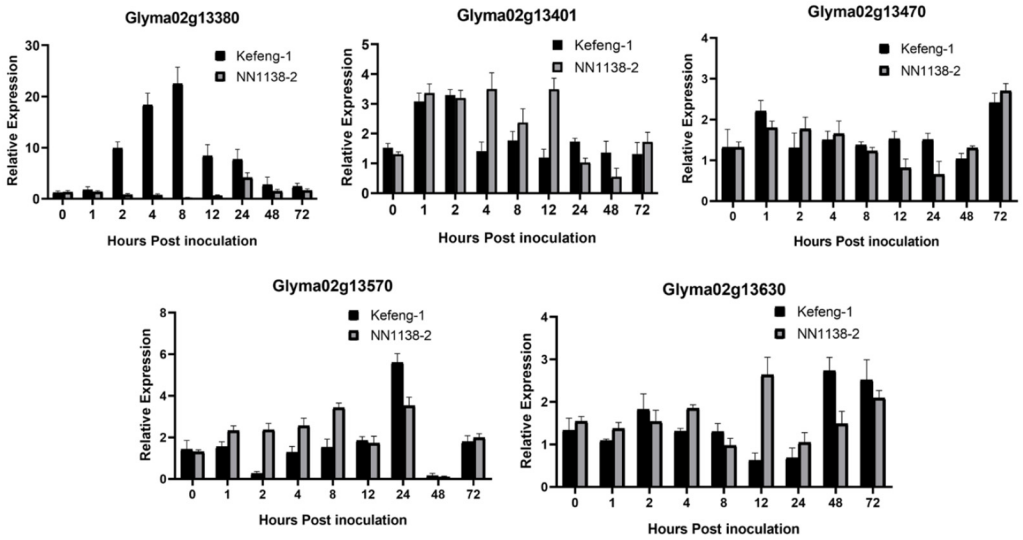


FIGURE 3
Differential expression profile of candidate genes resistant to SC4. Y-axes represent a relative expression of the genes between samples inoculated SMV and 0.01mol/L of phosphate buffer saline (PBS), and X-axes represent the time scale of sample collection.

not only expressed in the resistant parent but also in the susceptible parent at most time points; a higher upregulation was observed at 12 hpi in the susceptible parent and at 48 and 72 hpi in the resistant parent.

Similarly, the expression of six genes from the SC20 resistance region was observed (Figure 4). *Glyma02g13230* was upregulated in both parents at almost every time point, especially at 4, 12, and 48 hpi in the resistant parent and 4 and 12 hpi in the susceptible parent. *Glyma02g13361* and *Glyma02g13371* were expressed at low levels in both parents at 1, 4, 8, and 72 hpi. *Glyma02g13380* was upregulated in the resistant parent with a maximum magnitude of 8.24 folds at 8 hpi followed by 5.65 folds at 2 hpi and 3.18 folds at 48 hpi, whereas it was downregulated in the susceptible parent at 1, 2, 4, 8, and 48 hpi, with little upregulation at 72 hpi (Figure 4). *Glyma02g13401* was upregulated at 2 hpi (6.20 folds), followed by 1 hpi (3.79 folds) in the resistant parental line and at 4, 12, and 24 hpi in the susceptible parent. *Glyma02g13470* was upregulated only at 4 hpi (6.3) in the susceptible parents. Considering this, the expression results of SC4 and SC20 resistance genes, *Glyma02g13380* appears to be the most probable candidate gene for both SMV strains, as it is the

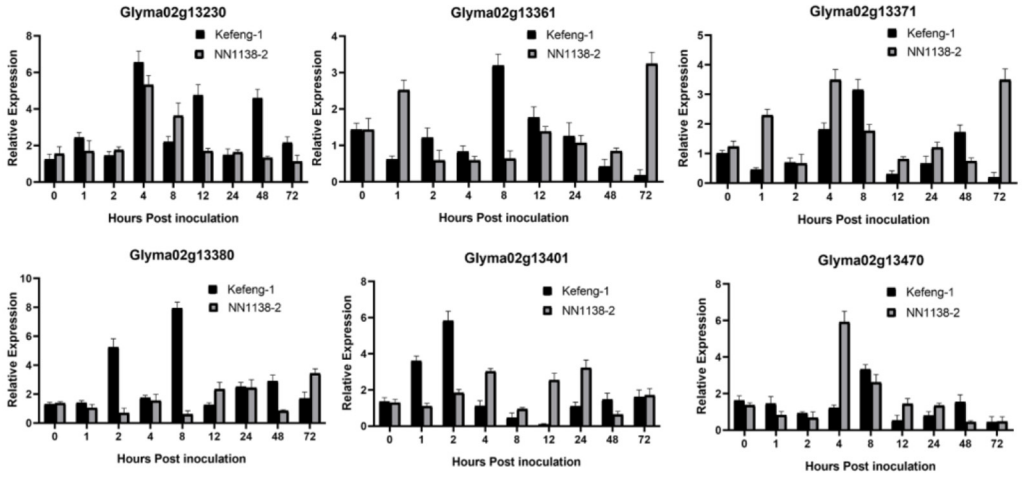


FIGURE 4
Differential expression profile of candidate genes resistant to SC20. Y-axes represent the relative expression of the genes between samples inoculated with SMV and 0.01mol/L of phosphate buffer saline (PBS), and X-axes represent the time scale of sample collection.

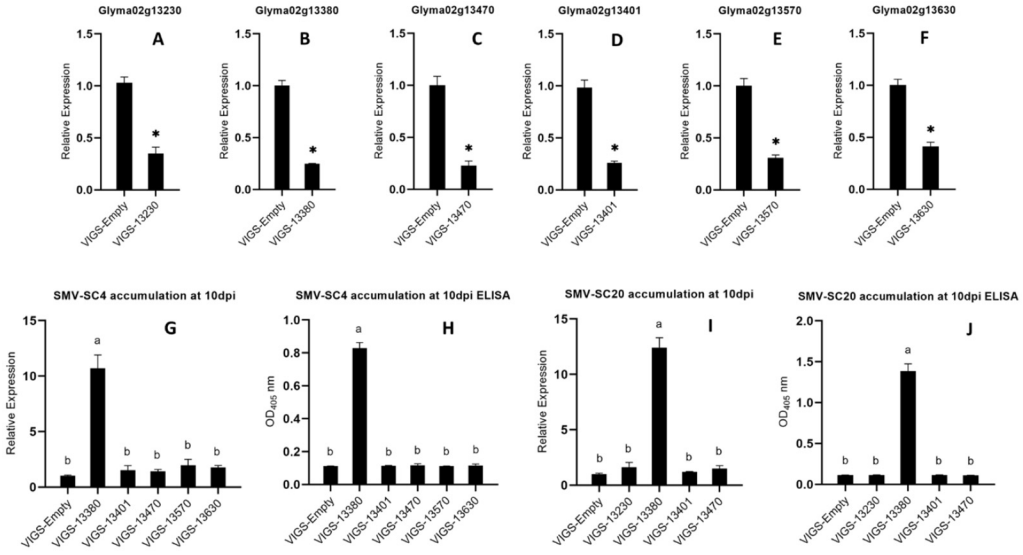


FIGURE 5
(A–F) Silencing efficiency of *Glyma02g13230, Glyma02g13380, Glyma02g13401, Glyma02g13470, Glyma02g13570*, and *Glyma02g13630* compared to that od control estimated by qRT-PCR. Bars indicating the standard errors, and Significant differences were tested by Student's t-test, $*P < 0.05$. **(G–J)** Accumulation of SMV strains SC4 and SC20 on the silenced plants ten days after inoculation detected by qRT-PCR and ELISA. G accumulation of SC4 detected by qRT-PCR, H accumulation of SC4 detected by ELISA, I accumulation of SC20 detected by qRT-PCR, and J accumulation of SC20 detected by ELISA. Different letters indicate the significant differences as tested by ANOVA/LSD ($P < 0.05$).

onlygene that shows significant and maximum upregulation in Kefeng-1 (resistant parent) and downregulation in NN1138-2 (susceptible parent) (Figures 3, 4).

3.5 Validation of *Glyma02g13380* through VIGS analysis

We used BPMV-based VIGS vectors to validate the functions of six SC4 and SC20 candidate resistant genes (*Glyma02g13230, Glyma02g13380, Glyma02g13401, Glyma02g13470, Glyma02g13570*, and *Glyma02g13630*) in Kefeng-1. Silencing vectors were designated as VIGS-13230, VIGS-13380, VIGS-13401, VIGS-13470, VIGS-13570, and VIGS-13630, respectively, and BPMV empty vector (VIGS-empty) was used as control. These VIGS vectors were propagated on susceptible cultivar NN1138-2 plants. Unifoliate leaves of Kefeng-1 plants were inoculated with BPMV recombinant viral vectors. After one week, the trifoliate leaves of these VIGS vector-inoculated plants were re-inoculated with the SMV strain SC4. The silencing efficacy of all silenced plants was examined by qRT-PCR and was found to be 74%, 73%, 74%, 65%, 61%, and 58%, respectively (Figure 5). SMV accumulation was detected by qRT-PCR and a DAS-ELISA. Plants silenced with *Glyma02g13380* accumulated more SMV than control plants, whereas those silenced with *Glyma02g13230, Glyma02g13401, Glyma02g13470, Glyma02g13570*, and *Glyma02g13630* did not show any deposits of SMV (Figure 5). Therefore, we can conclude that all six candidate genes were effectively silenced, and *Glyma02g13380* could be the most likely gene mediating resistance against SMV strains SC4 and SC20 in Kefeng-1.

3.6 *Glyma02g13380* in Kefeng-1 was distinct from the genes in Dabaima and Qihuang-1

Our candidate genes for SC4 and SC20 resistance in Kefeng-1 appeared to be dissimilar to the previously identified candidate resistance genes in Dabaima and Qihuang-1. However, to confirm this, we used the cloned coding sequences of SC4 and SC20 resistance candidates described by Wang et al. (2011) and Karthikeyan et al. (2018) and compared them with our candidate genes (Supplementary Figure 1). We also drafted models of the candidate genes according to the SoyKB (http://soykb.org). *Glyma02g13380* had a single exon of 1068 bp whereas the coding sequences of the other candidates were divided into several exons of different lengths (Figure 6A). The 2538 bp coding sequence of *Glyma14G38560* had seven exons (Figure 6B). *Glyma13g194700* and *Glyma13g195100* both exhibited a 3288-bp sequence and were divided into four exons (Figures 6C, D).

From the sequence and gene model comparisons, it is clear that *Glyma02g13380*, which is resistant to SC4 and SC20 in Kefeng-1, identified in the current study, was a completely different gene from the already identified candidate genes for these strains in Dabaima and Qihuang-1, and no evidence of translocation or mutation was found. As a result, it can be concluded that *Glyma02g13380*, a joint candidate gene for SC4 and SC20 resistance in Kefeng-1, is not a result of translocation or mutation of an already identified candidate but can be characterized as a distinct gene. Therefore, the two Mendelian genes on Chr.2 in Kefeng-1 for SC4 and SC20 resistance appear to represent a unique common gene, different

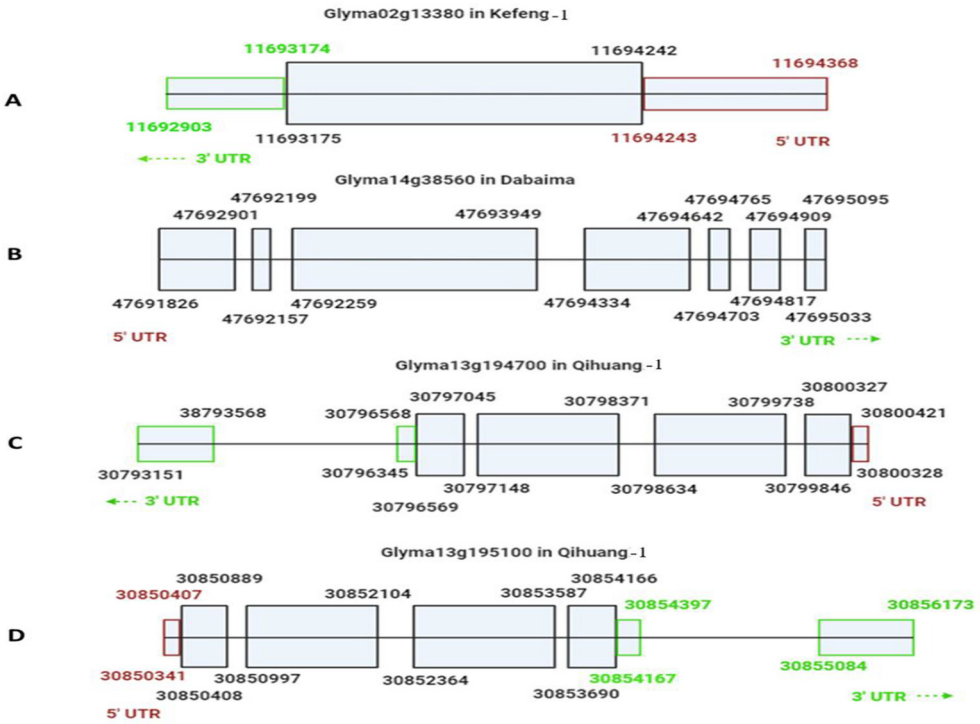


FIGURE 6
Drafted models of candidate resistant genes for SC4 and SC20 in different Chinese soybean cultivars. **(A)** *Glyma02g13380* responsible for SC4 resistance in Kefeng-1 showed one exon from 11693175 bp to 11694242 bp, 3'UTR from 11692903 bp to 11693174 bp (green color), and 5'UTR 11694243 bp to 11694368 bp (purple color). **(B)** *Glyma14g38560* was responsible for SC4 resistance in Dabaima, as determined by Wang et al. (2011) with a long coding sequence 47691826 bp-47695095 bp divided into seven exons of different sizes. **(C)** *Glyma13g194700* and **(D)** *Glyma13g195100* for SC20 in Qihuang-1 determined by Karthikeyan et al. (2018) with coding sequences of 30796569 bp–30800327 bp and 30850408 bp–30854166 bp, respectively, each divided into four exons.

from R_{SC4} in Dabaima and R_{SC20} in Qihuang-1. Furthermore, two more genes might be involved in a single strain resistance.

4 Discussion

4.1 Genetic mechanism of two separate genes conferring the resistance to the same strain

In Mendelian inheritance, a type of interaction is known between two genes, designated as duplicate genes. Hence, a trait conferred by the two genes interacted with their phenotypic segregation ratio in F_2 as 15 dominants: 1 recessive. This type of inheritance was explained by the two genes being duplicated, with their phenotypic expression being dominant for either one or two dominant alleles in their genetic construction. In the present study, we identified two genes, one on Chr.14 in Dabaima and one on Chr.2 in Kefeng-1, conferring the same resistance to SC4; however, the sequences of the two genes were different. Similar results were obtained for SC20. In both cases, the two independent genes were not identical. We are unsure whether this was an inheritance of duplicate genes. To explain this, we need to determine whether any particular sequence with both genes on Chr.14 and Chr.2 is responsible for SC4 resistance to conduct an inheritance study. The same is true for SC20. Alternatively, we need to find a reasonable hypothesis for duplicate genes, whether they may be different but exhibit the same function, or whether they should be strictly duplicated. Therefore, the present results do not describe a pair of duplicate genes. However, as mentioned earlier, this requires further study.

4.2 An efficient procedure in fine-mapping SMV resistance gene through SNPLDB genome-markers integrated with WGRS gene-markers

In fine mapping, a gene, usually derived from a subpopulation such as the residual heterozygote-derived population, was used. The preparation of the derived subpopulation is time-consuming. Taking advantage of genome-wide sequencing, linkage maps have been constructed using binary markers that cover the whole genome to locate genomic regions associated with specific SMV strain resistance. A genetic linkage map with high-resolution visibility is a powerful tool for identifying genomic regions with specific traits (Karthikeyan et al., 2017). In this study, two linkage maps were constructed using SNPLDB genome markers integrated with WGRS markers for each resistance gene. First, 3683 SNPLDB markers were used to build a high-density resolution map covering all 20 chromosomes of *Glycine max* to preliminarily map the SC4 and SC20 resistance genes. Second,

the WGRS gene marker data were used to further locate SNP(s) in the coding region of the resistance gene in both strains. These are efficient techniques for fine-mapping candidate gene(s) against specific SMV strains. The linkage map of 3683 SNPLDB markers built on 427 RILs revealed that the genomic region of 253 kb size affiliated with SC4 resistance was located between SNPLDB320-SNPLDB321 and the 375 kb genomic region associated with SC20 on Chr.2 (LGD1b), whereas the WGRS gene-marker map located the SC4 and SC20 loci between SNPs 11692903 and 14449821 with a shorter distance to 11692903.

The candidate genomic regions for SC4 and SC20 resistance that we uncovered on Chr.2 in the current study resembled many other reported intervals on Chr.2 and Chr.13 for SMV resistance. Zhao et al. (2016) discovered a 30.8-kb candidate resistance region on Chr.2 for R_{SC8}, and Li et al. (2015) discovered an 80-kb candidate resistance region on Chr.2 for R_{SC18} in the cultivar Kefeng-1. Karthikeyan et al. (2017) found a 500-kb region for R_{SC5} on Chr.2 in the cultivar Kefeng-1. Similarly, on Chr.13, possible resistance regions for specific SMV strains have been identified in many soybean cultivars, such as Qihuang-1, Suweon97, and PI96983 (Karthikeyan et al., 2018; Ma et al., 2016; Yang et al., 2013). These experiments and results can serve as supportive indices for the relevance of WGRS of the SNPLDB and gene marker mapping techniques for the determination of candidate resistant genes using a population of 427 recombinant inbred lines.

4.3 Functional validation of disease-resistant candidate genes via BPMV VIGS Vectors

Vectors based on bean pod mottle virus (BPMV) have been used extensively for functional studies of soybean genes (Zhang et al., 2010). After identifying and characterizing genes for specific traits, it is necessary to validate their functions. VIGS is a fast, efficient, and low-cost post-transcriptional gene silencing (PTGS) technique for investigating the functions of target genes (Senthil and Mysore, 2011). In the past, most VIGS vectors used for gene silencing were RNA-based viruses that have been used to silence many types of host crops, including (hot pepper, sweet pepper, Arabidopsis, Nicotiana, cotton, rice, barley, and corn) to remove intrinsic transcripts (Kant and Dasgupta, 2017). Traditional VIGS systems have been effectively employed to explore how plants respond to biotic and abiotic stressors (Shi et al., 2021). Boevink et al. (2016) employed VIGS to determine the mode of action of a novel fungal effector chemical in *Phytophthora infestans* (the pathogen responsible for late blight disease in the Solanaceae family) (Boevink et al., 2016). Keeping in mind the aforementioned successful functional validation of disease resistance genes, we also validated our candidate gene resistant to SMV in the soybean cultivar Kefeng-1 using the BPMV vector via virus-induced gene silencing and confirmed that *Glyma02g13380* is the most likely gene conferring resistance against SC4 and SC20 in Kefeng-1.

4.4 Non-nucleotide binding site-leucine-rich repeats (NBS-LLR) also confer resistance against SMV

Resistance to SMV in soybean is very complex, varies from cultivar to cultivar and strain to strain (Hajimorad et al., 2018), and is controlled by different genes belonging to different families. Mostly, genes belonging to the NBS-LRR gene family confer resistance against SMV and other diseases (Cao et al., 2020; Ji et al., 2020; Song et al., 2019; Wei et al., 2023). However, non-NBS-LRRs and unknown genes from different families have also been found to confer resistance to SMV (Khatabi et al., 2012; Wang et al., 2011). For example, one of the SC4 candidate resistance genes, *Glyma14g38580* in Dabaima (Wang et al., 2011) belongs to the P450 gene family, which is known to play multiple roles, such as regulation of growth and development processes, biosynthesis of secondary metabolites, and early defense against diseases in plants (Vasav and Barvkar, 2019). Similarly, two SMV isolates of strain G2 were highly similar in their coding sequences but differed in their virulence against SC4 resistance. This difference in virulence can be attributed to a single amino acid change in the P3 protein of the non-LLR gene (Khatabi et al., 2012). Almost all candidate genes conferring resistance to *Rsv4* on Chr.2 in soybeans are non-NBS-LRRs (Ishibashi et al., 2019). *Glyma02g13380* candidate resistance gene against SMV strains SC4 and SC20 in the current study is also uncharacterized and considered a non-NBS-LRR gene in soybean (https://www.soybase.org). Moreover, the homolog of this gene (AT1G69160) in *Arabidopsis thaliana* is involved in auxin transport (GO:0060918) and positive regulation of the auxin-mediated signaling pathway (GO:0010929). Auxins move into, out of, or within a cell, or between cells, with the help of mediators, such as transporters or pores. The auxin-mediated signaling pathway involves a series of molecular signals generated in response to auxin detection. Auxins are chiefly involved in plant growth and development, but their involvement in defense responses has also been suggested. The function of auxin incentive genes under biotic stress conditions is regulated by their differential expression in rice (Ghanashyam and Jain, 2009).

5 Conclusion

We identified and fine-mapped *Glyma02g13380*, a new single SC4 and SC20 resistance gene, on chromosome 2 in Kefeng-1 using GBS and SNPLDB fine-mapping approaches. The inheritance results confirmed that the resistance to SC4 and SC20 in Kefeng-1 was controlled by a single dominant gene. The function of this gene was validated using qRT-PCR, VIGS, and ELISA. The qRT-PCR and ELISA results confirmed that SMV was deposited only on plants that were silenced with *Glyma02g13380* but not on the control and other plants. Thus, *Glyma02g13380* was identified as a new candidate resistance gene against SC4 and SC20 in Kefeng-1.

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

MMR: Writing – original draft, Writing – review & editing. HJ: Data curation, Investigation, Writing – review & editing. MKR: Formal analysis, Writing – review & editing. BL: Data curation, Writing – review & editing. KL: Conceptualization, Supervision, Writing – review & editing. JG: Conceptualization, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. Zhongshan Biological Breeding Laboratory (ZSBBL-KY2023-03), China Agriculture Research System of MOF and MARA (No. CARS-04), Jiangsu Collaborative Innovation Center for Modern Crop Production (JCIC-MCP), and Collaborative Innovation Center for Modern Crop Production co-sponsored by Province and Ministry (CIC-MCP).

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.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2024.1518829/full#supplementary-material

SUPPLEMENTARY TABLE 1
Responses of 10 differential hosts to Soybean Mosaic Virus strains SC4 and SC20

SUPPLEMENTARY TABLE 2
qRT-PCR primers list with sequences, size and annealing temperature

SUPPLEMENTARY TABLE 3
Primer with their sequences for gene fragments confirmation, silencing efficiency and CP contents.

SUPPLEMENTARY TABLE 4
Functional annotation of seventeen candidate genes in genomic region related to SC4 and SC20 resistance

SUPPLEMENTARY FIGURE 1
Sequence alignment of resistant candidates for SC4 and SC20. *Glyma02g13380* in Kefeng-1 (this work), *Glyma14g38560* in Dabaima, and *Glyma13g194700* and *Glyma13g195100* in Qihuang-1.

References

Akond, M., Liu, S., Schoener, L., Anderson, J. A., Kantartzi, S. K., Meksem, K., et al. (2013). A SNP-based genetic linkage map of soybean using the SoySNP6K Illumina Infinium BeadChip genotyping array. *Plant Genet. Genomics Biotechnol.* 1, 80–89. doi: 10.5147/pggb.v1i3.154

Boevink, P. C., Wang, X., McLellan, H., He, Q., Naqvi, S., Armstrong, M. R., et al. (2016). A Phytophthora infestans RXLR effector targets plant PP1c isoforms that promote late blight disease. *Nat. Commun.* 7, 10311. doi: 10.1038/ncomms10311

Cao, Y., Liu, M., Long, H., Zhao, Q., Jiang, L., and Zhang, L. (2020). Hidden in plain sight: Systematic investigation of Leucine-rich repeat containing genes unveil the their regulatory network in response to Fusarium wilt in tung tree. *Int. J. Biol. Macromol.* 163, 1759–1767. doi: 10.1016/j.ijbiomac.2020.09.106

Cavanagh, C. R., Chao, S., Wang, S., Huang, B. E., Stephen, S., Kiani, S., et al. (2013). Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. *Proc. Natl. Acad. Sci.* 110, 8057–8062. doi: 10.1073/pnas.1217133110

Che, X. Y., Jiang, X., Liu, X. L., Luan, X. Y., Liu, Q., Cheng, X. F., et al. (2020). First report of Alfalfa mosaic virus on soybean in Heilongjiang, China. *Plant Dis.* 104, 3085–3085. doi: 10.1094/PDIS-04-20-0850-PDN

Chen, P., and Choi, C. (2007). "Characterization of genetic interaction between soybean and soybean mosaic virus," in *Molecular diagnosis of plant viruses*. Ed. G. P. Rao (Studium Press, Houston), 389–422.

Cho, E. K., and Goodman, R. M. (1982). Evaluation of resistance in soybeans to soybean mosaic virus strains 1. *Crop Sci.* 22, 1133–1136. doi: 10.2135/cropsci1982.0011183X002200060012x

Eid, M. A., Momeh, G. N., El-Shanshoury, A. E. R. R., Allam, N. G., and Gaafar, R. M. (2023). Comprehensive analysis of soybean cultivars' response to SMV infection: genotypic association, molecular characterization, and defense gene expressions. *J. Genet. Eng. Biotechnol.* 21, 102. doi: 10.1186/s43141-023-00558-x

Farnir, F., Coppieters, W., Arranz, J.-J., Berzi, P., Cambisano, N., Grisart, B., et al. (2000). Extensive genome-wide linkage disequilibrium in cattle. *Genome Res.* 10, 220–227. doi: 10.1101/gr.10.2.220

Fu, S., Zhan, Y., Zhi, H., Gai, J., and Yu, D. (2006). Mapping of SMV resistance gene Rsc-7 by SSR markers in soybean. *Genetica* 128, 63–69. doi: 10.1007/s10709-005-5535-9

Ganal, M. W., Altmann, T., and Röder, M. S. (2009). SNP identification in crop plants. *Curr. Opin. Plant Biol.* 12, 211–217. doi: 10.1016/j.pbi.2008.12.009

Ganal, M. W., Durstewitz, G., Polley, A., Bérard, A., Buckler, E. S., Charcosset, A., et al. (2011). A large maize (Zea mays L.) SNP genotyping array: development and germplasm genotyping, and genetic mapping to compare with the B73 reference genome. *PloS One* 6, e28334. doi: 10.1371/journal.pone.0028334

Ghanashyam, C., and Jain, M. (2009). Role of auxin-responsive genes in biotic stress responses. *Plant Signal. Behav.* 4, 846–848. doi: 10.4161/psb.4.9.9376

Hajimorad, M. R., Domier, L. L., Tolin, S. A., Whitham, S. A., and Saghai Maroof, M. (2018). *Soybean mosaic virus* : a successful potyvirus with a wide distribution but restricted natural host range. *A. Mol. Plant Pathol.* 19, 1563–1579. doi: 10.1111/mpp.12644

Hayes, A. J., Jeong, S. C., Gore, M. A., Yu, Y. G., Buss, G. R., Tolin, S. A., et al. (2004). Recombination within a nucleotide-binding-site/leucine-rich-repeat gene cluster produces new variants conditioning resistance to soybean mosaic virus in soybeans. *Genetics* 166, 493–503. doi: 10.1534/genetics.166.1.493

He, J., Meng, S., Zhao, T., Xing, G., Yang, S., Li, Y., et al. (2017). An innovative procedure of genome-wide association analysis fits studies on germplasm population and plant breeding. *Theor. Appl. Genet.* 130, 2327–2343. doi: 10.1007/s00122-017-2962-9

Hill, J. H., Soybean (2003). *Virus and virus-like diseases of major crops in developing countries* (Dordrecht: Springer Netherlands), 377–395. doi: 10.1007/978-94-007-0791-7_15

Hill, J. H., and Whitham, S. A. (2014). "Control of virus diseases in soybeans," in *Control of plant virus diseases*. Eds. G. Loebenstein and N. B. T. Katis (San Diego, CA, USA: Academic Press), 355–390. doi: 10.1016/B978-0-12-801246-8.00007-X

Hwang, T. Y., Moon, J. K., Yu, S., Yang, K., Mohankumar, S., Yu, Y. H., et al. (2006). Application of comparative genomics in developing molecular markers tightly linked to the virus resistance gene Rsv 4 in soybean. *Genome* 49, 380–388. doi: 10.1139/g05-111

Ishibashi, K., Saruta, M., Shimizu, T., Shu, M., Anai, T., Komatsu, K., et al. (2019). Soybean antiviral immunity conferred by dsRNase targets the viral replication complex. *Nature Commun.* 10, 4033. doi: 10.1038/s41467-019-12052-5

Jeong, S. C., and Saghai Maroof, M. A. (2004). Detection and genotyping of SNPs tightly linked to two disease resistance loci, *Rsv1* and *Rsv3*, of soybean. *Plant Breed.* 123, 305–310. doi: 10.1111/j.1439-0523.2004.00981.x

Ji, C., Ji, Z., Liu, B., Cheng, H., Liu, H., Liu, S., et al. (2020). Xa1 allelic R genes activate rice blight resistance suppressed by interfering TAL effectors. *Plant Commun.* 1, 100087. doi: 10.1016/j.xplc.2020.100087

Jones, E., Chu, W.-C., Ayele, M., Ho, J., Bruggeman, E., Yourstone, K., et al. (2009). Development of single nucleotide polymorphism (SNP) markers for use in commercial maize (Zea mays L.) germplasm. *Mol. Breed.* 24, 165–176. doi: 10.1007/s11032-009-9281-z

Kant, R., and Dasgupta, I. (2017). Phenotyping of VIGS-mediated gene silencing in rice using a vector derived from a DNA virus. *Plant Cell Rep.* 36, 1159–1170. doi: 10.1007/s00299-017-2156-6

Karthikeyan, A., Li, K., Jiang, H., Ren, R., Li, C., Zhi, H., et al. (2017). Inheritance, fine-mapping, and candidate gene analyses of resistance to soybean mosaic virus strain SC5 in soybean. *Mol. Genet. Genomics* 292, 811–822. doi: 10.1007/s00438-017-1310-8

Karthikeyan, A., Li, K., Li, C., Yin, J., Li, N., Yang, Y., et al. (2018). Fine-mapping and identifying candidate genes conferring resistance to Soybean mosaic virus strain SC20 in soybean. *Theor. Appl. Genet.* 131, 461–476. doi: 10.1007/s00122-017-3014-1

Khatabi, B., Fajolu, O. L., Wen, R. H., and Hajimorad, M. R. (2012). Evaluation of American isolates of soybean mosaic virus for gain of virulence on genotype soybeans with special emphasis on resistance-breaking determinants on Rsv4. *Mol. Plant Pathol.* 13, 1077–1088. doi: 10.1111/j.1364-3703.2012.00817.x

Kosambi, D. D. (2016). "The estimation of map distances from recombination values,". Ed. D. D. Kosambi: *selected works in mathematics and statistics*, (Springer India, New Delhi), 125–130. doi: 10.1007/978-81-322-3676-4_16

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324

Li, Y., Liu, X., Deng, W., Liu, J., Fang, Y., Liu, Y., et al. (2022). Fine mapping the soybean mosaic virus resistance gene in soybean cultivar heinong 84 and development of CAPS markers for rapid identification. *Viruses* 14, 2533. doi: 10.3390/v14112533

Li, K., Ren, R., Adhimoolam, K., Gao, L., Yuan, Y., Liu, Z., et al. (2015). Genetic analysis and identification of two soybean mosaic virus resistance genes in soybean. *Plant Breed.* 134, 684–695. doi: 10.1111/pbr.12315

Li, K., Yang, Q. H., Zhi, H. J., and Gai, J. Y. (2010). Identification and Distribution of *Soybean mosaic virus* Strains in Southern China. *Plant Dis.* 94, 351–357. doi: 10.1094/PDIS-94-3-0351

Li, R., Yu, C., Li, Y., Lam, T. W., Yiu, S. M., Kristiansen, K., et al. (2009). SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics* 25, 1966–1967. doi: 10.1093/bioinformatics/btp336

Li, H., Zhi, H., Gai, J., Guo, D., Wang, Y., Li, K., et al. (2006). Inheritance and gene mapping of resistance to soybean mosaic virus strain SC14 in soybean. *J. Integr. Plant Biol.* 48, 1466–1472. doi: 10.1111/j.1744-7909.2006.00365.x

Liu, J.-Z., Fang, Y., and Pang, H. (2016). The current status of the soybean-soybean mosaic virus (SMV) pathosystem. *Front. Microbiol.* 7. doi: 10.3389/fmicb.2016.01906

Liu, H., Zheng, H., Xiang, W., Song, Y., Li, B., Yin, J., et al. (2023). Construction and characterization of the infectious cDNA clone of the prevalent Chinese strain SC3 of soybean mosaic virus. *Phytopathol. Res.* 5, 9. doi: 10.1186/s42483-023-00164-2

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262

Ma, F. F., Wu, X. Y., Chen, Y. X., Liu, Y. N., Shao, Z. Q., Wu, P., et al. (2016). Fine mapping of the Rsv1-h gene in the soybean cultivar Suweon 97 that confers resistance to two Chinese strains of the soybean mosaic virus. *Theor. Appl. Genet.* 129, 2227–2236. doi: 10.1007/s00122-016-2769-0

McNally, K. L., Childs, K. L., Bohnert, R., Davidson, R. M., Zhao, K., Ulat, V. J., et al. (2009). Genomewide SNP variation reveals relationships among landraces and modern varieties of rice. *Proc. Natl. Acad. Sci.* 106, 12273–12278. doi: 10.1073/pnas.0900992106

Roberts, A., McMillan, L., Wang, W., Parker, J., Rusyn, I., and Threadgill, D. (2007). Inferring missing genotypes in large SNP panels using fast nearest-neighbor searches over sliding windows. *Bioinformatics* 23, i401–i407. doi: 10.1093/bioinformatics/btm220

Rostoks, N., Mudie, S., Cardle, L., Russell, J., Ramsay, L., Booth, A., et al. (2005). Genome-wide SNP discovery and linkage analysis in barley based on genes responsive to abiotic stress. *Mol. Genet. Genomics* 274, 515–527. doi: 10.1007/s00438-005-0046-z

Senthil, K. M., and Mysore, K. S. (2011). New dimensions for VIGS in plant functional genomics. *Trends Plant Sci.* 16, 656–665. doi: 10.1016/j.tplants.2011.08.006

Shi, G., Hao, M., Tian, B., Cao, G., Wei, F., and Xie, Z. (2021). A methodological advance of tobacco rattle virus-induced gene silencing for functional genomics in plants. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.671091

Song, H., Guo, Z., Hu, X., Qian, L., Miao, F., Zhang, X., et al. (2019). Evolutionary balance between LRR domain loss and young NBS–LRR genes production governs disease resistance in Arachis hypogaea cv. Tifrunner. *BMC Genomics* 20, 844. doi: 10.1186/s12864-019-6212-1

Song, Q., Hyten, D. L., Jia, G., Quigley, C. V., Fickus, E. W., Nelson, R. L., et al. (2013). Development and evaluation of soySNP50K, a high-density genotyping array for soybean. *PloS One* 8, e54985. doi: 10.1371/journal.pone.0054985

Takahashi, K., Tanaka, T., Iida, W., and Tsuda, Y. (1980). Studies on virus diseases and causal viruses of soybean in Japan. *Bull. Tohoku Natl. Agric. Exp. Station.* 62, 1–130.

Tian, H. L., Wang, F. G., Zhao, J. R., Yi, H. M., Wang, L., Wang, R., et al. (2015). Development of maizeSNP3072, a high-throughput compatible SNP array, for DNA fingerprinting identification of Chinese maize varieties. *Mol. Breed.* 35, 136. doi: 10.1007/s11032-015-0335-0

Vasav, A. P., and Barvkar, V. T. (2019). Phylogenomic analysis of cytochrome P450 multigene family and their differential expression analysis in tomato. suggested tissue specific promoters. *BMC Genomics* 20, 116. doi: 10.1186/s12864-019-5483-x

Wang, D., Ma, Y., Liu, N., Yang, Z., Zheng, G., and Zhi, H. (2011). Fine mapping and identification of the soybean R SC4 resistance candidate gene to soybean mosaic virus. *Plant Breed.* 130, 653–659. doi: 10.1111/j.1439-0523.2011.01888.x

Wang, D., Tian, Z., Li, K., Li, H., Huang, Z., Hu, G., et al. (2013). Identification and variation analysis of soybean mosaic virus strains in Shandong, Henan and Anhui provinces of China. *Soybean Sci.* 32, 806–809.

Wei, W., Wu, X., Garcia, A., McCoppin, N., Viana, J. P. G., Murad, P. S., et al. (2023). An NBS-LRR protein in the Rpp1 locus negates the dominance of Rpp1-mediated resistance against Phakopsora pachyrhizi in soybean. *Plant J.* 113, 915–933. doi: 10.1111/tpj.16038

Yang, N., Qiu, Y., Shen, Y., Xu, K., and Yin, J. (2024). Genome-wide analysis of soybean mosaic virus reveals diverse mechanisms in parasite-derived resistance. *Agronomy* 14, 1457. doi: 10.3390/agronomy14071457

Yang, Y., Zheng, G., Han, L., Dagang, W., Yang, X., Yuan, Y., et al. (2013). Genetic analysis and mapping of genes for resistance to multiple strains of Soybean mosaic virus in a single resistant soybean accession PI 96983. *Theor. Appl. Genet.* 126, 1783–1791. doi: 10.1007/s00122-013-2092-y

Zhan, Y., Zhi, H., Yu, D., and Gai, J. (2006). Identification and distribution of SMV strains in Huang-Huai Valleys. *Sci. Agric. Sin.* 39, 2009–2015.

Zhang, C., Bradshaw, J. D., Whitham, S. A., and Hill, J. H. (2010). The development of an efficient multipurpose bean pod mottle virus viral vector set for foreign gene expression and RNA silencing. *Plant Physiol.* 153, 52–65. doi: 10.1104/pp.109.151639

Zhang, Y., Song, J., Wang, L., Yang, M., Hu, K., Li, W., et al. (2022). Identifying quantitative trait loci and candidate genes conferring resistance to soybean mosaic virus SC7 by quantitative trait loci-sequencing in soybean. *Front. Plant Sci.* 13, 843633. doi: 10.3389/fpls.2022.843633

Zhang, C., Yang, C., Whitham, S. A., and Hill, J. H. (2009). Development and use of an efficient DNA-based viral gene silencing vector for soybean. *Mol. Plant-Microbe Interact.* 22, 123–131. doi: 10.1094/MPMI-22-2-0123

Zhao, L., Wang, D., Zhang, H., Shen, Y., Yang, Y., Li, K., et al. (2016). Fine mapping of the R SC8 locus and expression analysis of candidate SMV resistance genes in soybean. *Plant Breed.* 135, 701–706. doi: 10.1111/pbr.12428

Zhi, H., Gai, J., Chen, Y., Liao, Q., Guo, D., Wang, Y., et al. (2005). Evaluation of resistance to SMV of the entries in the national uniform soybean tests, (2002-2004). *Soybean Sci.* 24, 189–194.

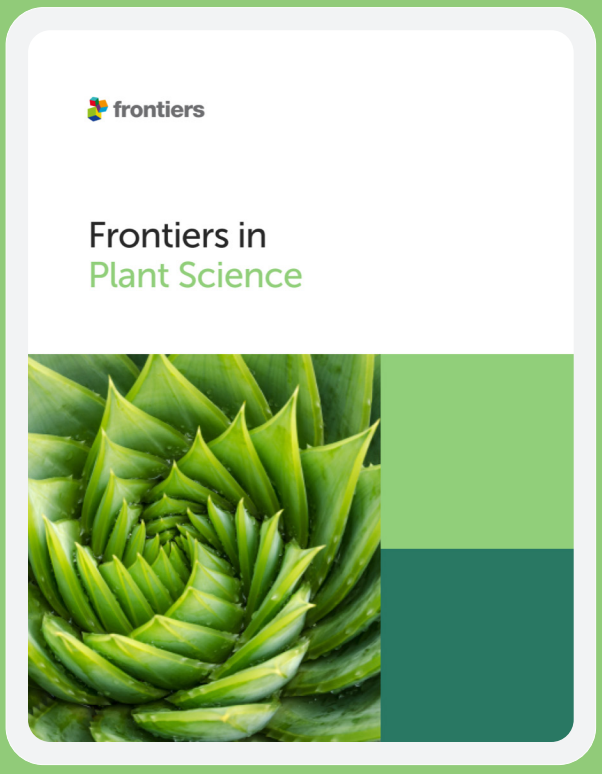
Frontiers in
Plant Science

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