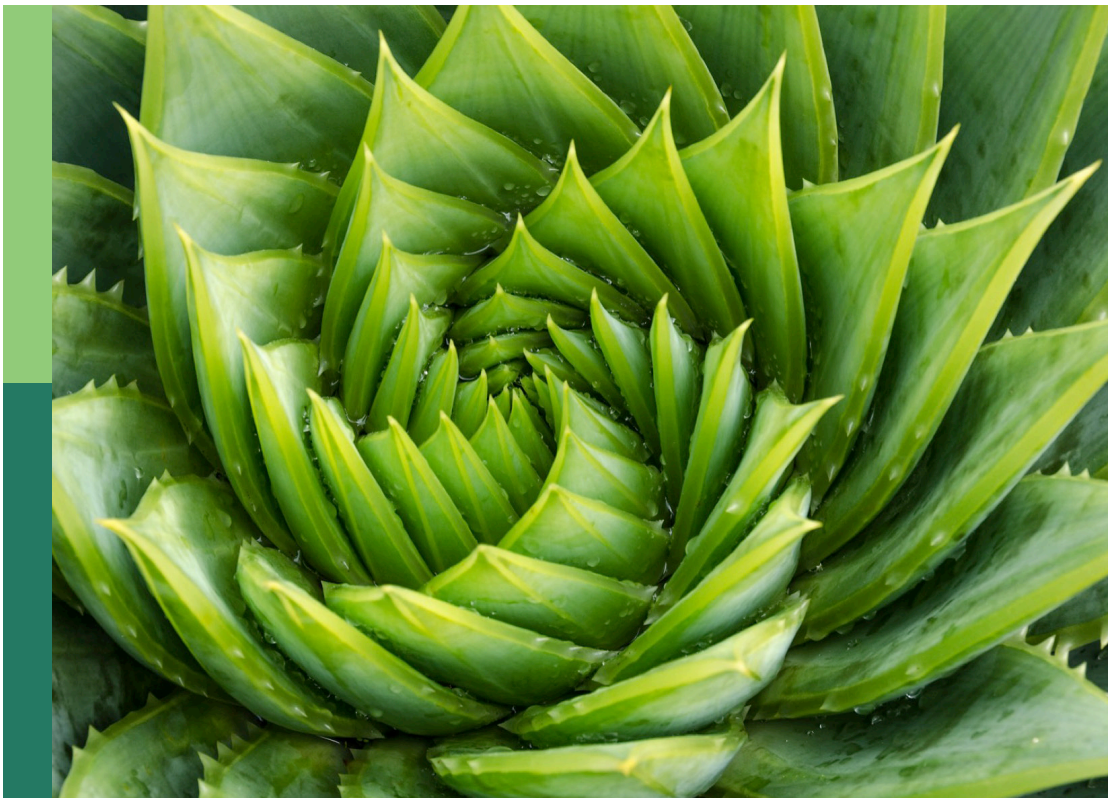
frontiers | Research Topics

Utilizing advanced genomics and biochemical tools to strengthen crop adaptation for biotic and abiotic stresses

Edited by
Gurjeet Singh, Santosh Gudi, Diaa Abd El Moneim and Simardeep Kaur

Published in
Frontiers in Plant Science



ISSN 1664-8714
ISBN 978-2-8325-7476-8
DOI 10.3389/978-2-8325-7476-8

Generative AI statement
Any alternative text (Alt text) provided alongside figures in the articles in this ebook has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

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

Utilizing advanced genomics and biochemical tools to strengthen crop adaptation for biotic and abiotic stresses

Topic editors

Gurjeet Singh — Texas A and M University, United States
Santosh Gudi — North Dakota State University, United States
Diaa Abd El Moneim — Arish University, Egypt
Simardeep Kaur — The ICAR Research Complex for North Eastern Hill Region (ICAR RC NEH), India

Citation

Singh, G., Gudi, S., Abd El Moneim, D., Kaur, S., eds. (2026). *Utilizing advanced genomics and biochemical tools to strengthen crop adaptation for biotic and abiotic stresses*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-7476-8

Table of contents

frontiers | Frontiers in Plant Science

TYPE Editorial
PUBLISHED 26 January 2026
DOI 10.3389/fpls.2025.1766137

OPEN ACCESS

EDITED AND REVIEWED BY
Jian You Wang,
Academia Sinica, Taiwan

*CORRESPONDENCE
Gurjeet Singh
gurjeet-pbg@pau.edu

RECEIVED 12 December 2025
ACCEPTED 22 December 2025
PUBLISHED 26 January 2026

CITATION
Singh G, Gudi S, Kaur S and Abd El Moneim D (2026) Editorial: Utilizing advanced genomics and biochemical tools to strengthen crop adaptation for biotic and abiotic stresses. *Front. Plant Sci.* 16:1766137. doi: 10.3389/fpls.2025.1766137

Editorial: Utilizing advanced genomics and biochemical tools to strengthen crop adaptation for biotic and abiotic stresses

Gurjeet Singh[1]*, Santosh Gudi[2], Simardeep Kaur[3,4] and Diaa Abd El Moneim[5]

[1]Texas A&M University, AgriLife Research Center, Beaumont, TX, United States, [2]North Dakota State University, Fargo, ND, United States, [3]Division of Crop Science, ICAR- Research Complex for North Eastern Hill Region, Meghalaya, Umiam, India, [4]Department of Plant Sciences, North Dakota State University, Fargo, ND, United States, [5]Department of Plant Production (Genetic Branch), Faculty of Environmental Agricultural Sciences, Arish University, El-Arish, Egypt

KEYWORDS

climate change, genomics tools, multi-omics approaches, biochemical pathways, biotic stress resistance, abiotic stress tolerance, sustainability

Editorial on the Research Topic
Utilizing advanced genomics and biochemical tools to strengthen crop adaptation for biotic and abiotic stresses

Overview

Global agriculture faces unprecedented challenges such as emerging pests and diseases, and increasing incidence of abiotic stresses such as drought, salinity, waterlogging, and heat, threatening crop productivity and food security. The Research Topic "*Utilizing Advanced Genomics and Biochemical Tools to Strengthen Crop Adaptation for Biotic and Abiotic Stresses,*" published in *Frontiers in Plant Science*, comprises 11 scientific articles (ten original research and one review) contributed by 86 researchers worldwide. This Research Topic showcases the integration of advanced genomics, molecular breeding, and biochemical approaches to develop climate-resilient crops. Using high-throughput sequencing, genome-wide association studies (GWAS), QTL mapping, transcriptome and metabolomic analyses, key genomic regions and metabolic pathways linked to stress tolerance were identified in wheat, maize, soybean, sugarcane, pigeonpea, *Brassica napus*, and *Salvia miltiorrhiza*. Significant contributions include the identification of stripe rust resistance loci in wheat, drought-tolerant clones in sugarcane, and pod borer tolerance genetic loci in pigeonpea. Machine learning approaches prioritized stress-associated genes in maize, while comparative transcriptomics in *S. miltiorrhiza* revealed MAPK cascade roles in pathogen response and metabolite accumulation. Soybean studies identified genes and pathways that enhance tolerance to drought, salinity, and biotic stresses, including the CAM/CML gene family, which confers dual viral and fungal resistance. Additional insights include drought-responsive microRNAs in wheat and waterlogging resilience in *B. napus*. A review of genotyping in climate-smart breeding emphasizes the use of integrative tools to accelerate genetic improvement. Therefore, this RT demonstrates how integrating

genomics, biochemical profiling, and computational biology can accelerate the development of high-yielding, stress-resilient cultivars as well as helps in contributing to sustainable, climate-smart agriculture.

Published articles and summaries

1. *Identification of new genetic resources for drought tolerance-related traits from the world Erianthus germplasm collection* (Ramanathan et al.)

A diverse panel of 223 *Erianthus* germplasm accessions from seven countries were evaluated under field-imposed drought stress to identify donor clones for sugarcane improvement. Physiological screening and multi-year phenotyping enabled the development of a 91-clone drought response association panel. GWAS using 1,044 high-quality SNPs identified 43 QTNs (quantitative trait nucleotides) associated with drought-adaptive traits. Candidate genes analysis from the identified QTL regions revealed the set of genes involved in stress perception and signaling, including TOR2, TMK1, potassium, and nitrate transporters. These drought-tolerant clones, QTNs, and gene targets provide valuable resources for developing drought-resilient sugarcane cultivars through marker-assisted breeding.

2. *Genome wide association studies (GWAS) for identification of stripe rust resistance loci in diverse wheat genotypes* (Tanwar et al.)

Stripe rust, caused by *Puccinia striiformis f.* sp. *tritici* (Pst), remains a major threat of wheat productivity in North India due to rapidly evolving virulent races. To identify durable resistance, a GWAS was conducted on 652 elite wheat genotypes using 1,938 DArTseq SNPs markers and phenotypic data from four locations. The analysis revealed 27 significant genomic regions associated with stripe rust resistance, including loci on chromosomes 2B, 6A, and 6B. Candidate gene analysis from the QTL region identified defense-related genes such as NBLRR, F-box, LRR, and kinase families. These loci provide strong targets for developing user-friendly markers and accelerating breeding of rust-resistant wheat varieties.

3. *Genome-wide characterization and stress-responsive expression analysis of the cinnamoyl-CoA reductase gene family in soybean* (Li et al.)

A genome-wide analysis identified 15 CCR (cinnamoyl-CoA reductase) members in soybean across 12 chromosomes using comparative genomics, domain validation, and phylogenetics. Promoter, motif, and synteny analyses revealed diversified regulatory elements and evolutionary expansion. Transcriptome profiling under four abiotic stresses (salt, alkaline, drought, and osmotic) showed strong root-specific and stress-responsive expression, with *GmCCR1*, *GmCCR4*, *GmCCR7*, *GmCCR8*, and *GmCCR15* significantly upregulated. *GmCCR4* exhibited the most robust induction, thus indicating a key role in lignin-mediated stress adaptation. These CCR members provide molecular targets for marker-assisted breeding and genetic engineering to develop salt-alkali-tolerant, climate-resilient soybean varieties, supporting sustainable production in degraded soils.

4. *Dissecting genomic regions and candidate genes for pod borer resistance and component traits in pigeonpea minicore collection* (Moghiya et al.)

A panel of 146 pigeonpea minicore accessions, along with resistant and susceptible checks, were evaluated over three field seasons for pod borer resistance and related traits. Whole-genome resequencing generated 499,980 SNPs, enabling multi-locus GWAS using SUPER and FarmCPU models. Analysis identified 14 significant MTAs across five chromosomes, linked to key candidate genes including *carboxylesterase 15*, *microtubule-associated protein 5*, and *FAR1-related sequence*. Four accessions (ICP 10503, ICP 655, ICP 9691, and ICP 9655) showed moderate resistance in pod bioassays. These MTAs, genes, and resistant lines provide strong resources for marker-assisted breeding of pod borer–resistant pigeonpea varieties.

5. *Comprehensively characterize the soybean CAM/CML gene family, as it provides resistance against both the soybean mosaic virus and* Cercospora sojina *pathogens* (Zhang et al.)

This study identifies key regulators of broad-spectrum disease resistance through a genome-wide characterization of 113 CAM/CML genes (11 *GmCAMs* and 102 *GmCMLs*) in soybean. Phylogenetic, structural, and cis-regulatory analyses revealed 14 evolutionary groups enriched with hormone and stress-responsive elements. Expression profiling under soybean mosaic virus and *Cercospora sojina* infections at the V3 stage, five susceptible and five tolerant genotypes were identified 19 CAM/CML genes responsive to both pathogens, with *GmCAM4*, *GmCML23*, and *GmCML47* strongly associated with resistance. These findings provide a valuable resource for understanding calcium-mediated defense mechanisms and offer promising targets for molecular breeding and genome editing to develop multi-disease-resistant soybean cultivars.

6. *Predictive prioritization of genes significantly associated with biotic and abiotic stresses in maize using machine learning algorithms* (Pradhan et al.)

This study integrated 39,756 RNA-seq datasets from maize under diverse biotic and abiotic stresses to identify candidate genes for stress resilience. Using seven machine learning-based models and WGCNA, 235 top-ranked genes were prioritized, including hub genes such as *Zm00001eb176680* (*bZIP* transcription factor 68), *Zm00001eb176940* (glycine-rich structural protein 2), and *Zm00001eb179190* (*ALDH11*). Promoter analysis revealed enrichment for abscisic acid and antioxidant-responsive elements, suggesting regulatory roles in stress adaptation. These findings provide a comprehensive resource for understanding maize stress responses and offer key molecular targets for functional validation, genome editing, and molecular breeding to develop multi-stress-resilient maize cultivars for sustainable agriculture.

7. *Novel and conserved drought-responsive microRNAs expression analysis in root tissues of wheat (Triticum asetivum L.) at reproductive stage* (Sharma et al.)

This study presents a comprehensive analysis of drought-responsive microRNAs (miRNAs) in drought-tolerant (NI5439) and susceptible (WL711) wheat genotypes under control and drought stress conditions at the booting stage. A total of 364 miRNAs (306 known and 58 novel) were identified from four

sRNA libraries. 18 miRNAs showed significant changes in expression after stress treatment, with 2,300 predicted target genes involved in signal transduction, epigenetic regulation, and development. Ten novel miRNAs were validated via qRT-PCR, confirming genotype-specific responses. These findings expand the catalog of drought-responsive miRNAs, providing molecular targets for functional genomics and offering future opportunities for genetic engineering and breeding drought-resilient wheat cultivars.

8. *Comparative transcriptomic analysis and genome-wide identification provide insights into the potential role of fungal-responsive MAPK cascade genes in tanshinone accumulation in Salvia miltiorrohiza* (Abozeid et al.)

This study provides the first genome-wide characterization of the MAPK gene family in *S. miltiorrhiza* and its role in tanshinone biosynthesis under fungal elicitation. A total of 17 *SmMAPKs*, 7 *SmMAPKKs*, and 22 *SmMAPKKKs* genes distributed across nine chromosomes were identified and were grouped into TEY and TDY subfamilies. Transcriptome and HPLC analyses revealed that yeast extract and *Aspergillus niger* significantly enhanced tanshinone accumulation, particularly cryptotanshinone and dihydrotanshinone. *SmMAPK4* and *SmMAPKK5* showed strong positive correlations with tanshinone content, suggesting a regulatory role. These findings provide molecular targets for genetic engineering and fungal elicitor-based strategies to enhance tanshinone production, paving the way for improved medicinal applications.

9. *Deciphering salt tolerance mechanisms in synthetic hexaploid and bread wheat under humic acid application: physiological and genetic perspectives* (Alghabari and Shah)

This study evaluated the effect of humic acid (HA) on salt tolerance in four synthetic hexaploid (SH) and three bread wheat (BW) genotypes for physiological, biochemical, and genetic traits. HA enhanced chlorophyll content (33.3-100%), photosynthesis (31.2-50%), and antioxidant enzyme activities (SOD, POD, CAT), while reduced Na^+/K^+ ratio (33.3-50%), proline (20-28.5%), and glycine betaine (42.8-77.7%) under salt stress. Salinity-associated genes (*TaNHX1*, *TaHKT1,4*, *TaAKT1*, *TaPRX2A*, *TaSOD*, and *TaCAT1*) were upregulated, whereas *TaP5CS* was downregulated, in SH wheat lines, indicating superior tolerance. SH genotypes can serve as a bridge to transfer salt tolerance traits into wheat breeding programs. Future studies should conduct field evaluations across diverse soils and climates to optimize HA application and assess long-term impacts on wheat salt resilience.

10. *Genome-wide identification of PDX and expression analysis under waterlogging stress exhibit stronger waterlogging tolerance in transgenic Brassica napus plants overexpressing the BnaPDX1.3 gene compared to wild-type plants* (Yao et al.)

This study characterizes the PDX gene family in *B. napus* variety G218, revealing 13 PDX genes with high evolutionary conservation and roles in waterlogging stress response. Functional analysis of *BnaPDX1.3* overexpressing plants demonstrated enhanced vitamin B6 synthesis, stronger antioxidant enzyme activity, stable ROS homeostasis, and improved biomass under waterlogging compared to wild-type plants. The results highlight *BnaPDX1.3* as a key regulator of waterlogging tolerance *B. napus*. These findings provide a foundation for molecular breeding and genetic engineering strategies to develop waterlogging-tolerant *B. napus* cultivars. Author concluded that future research could explore regulatory networks, cross-stress tolerance, and field-level validation to accelerate climate-resilient crop development.

11. *The importance of genotyping within the climate-smart plant breeding value chain integrative tools for genetic enhancement programs* (Garcia-Oliveira et al.)

This review highlights the key role of genotyping and integrative tools in climate-smart plant breeding to enhance crop resilience and productivity. Modern approaches, including genome editing, mutation breeding, multi-omics, microRNAs, and digital technologies, enable precise targeting of loci controlling complex traits and support real-time decision-making. Climate-smart plant breeding innovations are relevant for smallholder farmers across diverse agro-climatic zones, helping crops adapt to climate variability, nutritional demands, and consumer preferences. The article also emphasizes responsible adoption, addressing genetic erosion, biodiversity, and intellectual property concerns. Overall, integrating these tools offers a holistic strategy for sustainable, climate-resilient agricultural systems.

Author contributions

GS: Conceptualization, Investigation, Visualization, Writing – original draft, Writing – review & editing. SG: Conceptualization, Visualization, Writing – review & editing. SK: Conceptualization, Writing – review & editing. DM: Conceptualization, Writing – review & editing.

Conflict of interest

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

The authors DM and SK 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) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

frontiers | Frontiers in Plant Science

TYPE Review
PUBLISHED 06 February 2025
DOI 10.3389/fpls.2024.1518123

OPEN ACCESS

EDITED BY
Simardeep Kaur,
The ICAR Research Complex for North Eastern Hill Region (ICAR RC NEH), India

REVIEWED BY
Sonu Shekhawat,
Indian Agricultural Research Institute (ICAR), India
Deepesh Kumar,
Indian Council of Agricultural Research, India

*CORRESPONDENCE
Ana Luísa Garcia-Oliveira
a.oliveira@cgiar.org

RECEIVED 29 October 2024
ACCEPTED 25 November 2024
PUBLISHED 06 February 2025

CITATION
Garcia-Oliveira AL, Ortiz R, Sarsu F, Rasmussen SK, Agre P, Asfaw A, Kante M and Chander S (2025) The importance of genotyping within the climate-smart plant breeding value chain – integrative tools for genetic enhancement programs. *Front. Plant Sci.* 15:1518123. doi: 10.3389/fpls.2024.1518123

The importance of genotyping within the climate-smart plant breeding value chain – integrative tools for genetic enhancement programs

Ana Luísa Garcia-Oliveira [1]*, Rodomiro Ortiz [2], Fatma Sarsu [3], Søren K. Rasmussen [4], Paterne Agre [5], Asrat Asfaw [5], Moctar Kante [6] and Subhash Chander [7]

[1]Genetic Resources Program, Alliance Bioversity International and International Center for Tropical Agriculture (CIAT), Cali, Colombia, [2]Department of Plant Breeding, Swedish University of Agricultural Sciences, Alnarp, Sweden, [3]Plant Breeding and Genetics Section, Joint FAO/IAEA Center, International Atomic Energy Agency, Vienna, Austria, [4]Independent Researcher, Roskilde, Denmark, [5]Yam Breeding Unit, International Institute of Tropical Agriculture, Ibadan, Nigeria, [6]Genetics, Genomics, and Crop Improvement Division, International Potato Center, Lima, Peru, [7]Oilseeds Section, Department of Genetics & Plant Breeding, CCS Haryana Agricultural University, Hisar, India

The challenges faced by today's agronomists, plant breeders, and their managers encompass adapting sustainably to climate variability while working with limited budgets. Besides, managers are dealing with a multitude of issues with different organizations working on similar initiatives and projects, leading to a lack of a sustainable impact on smallholder farmers. To transform the current food systems as a more sustainable and resilient model efficient solutions are needed to deliver and convey results. Challenges such as logistics, labour, infrastructure, and equity, must be addressed alongside adapting to increasingly unstable climate conditions which affect the life cycle of transboundary pathogens and pests. In this context, transforming food systems go far beyond just farmers and plant breeders and it requires substantial contributions from industry, global finances, transportation, energy, education, and country developmental sectors including legislators. As a result, a holistic approach is essential for achieving sustainable and resilient food systems to sustain a global population anticipated to reach 9.7 billion by 2050 and 11.2 billion by 2100. As of 2021, nearly 193 million individuals were affected by food insecurity, 40 million more than in 2020. Meanwhile, the digital world is rapidly advancing with the digital economy estimated at about 20% of the global gross domestic product, suggesting that digital technologies are increasingly accessible even in areas affected by food insecurity. Leveraging these technologies can facilitate the development of climate-smart cultivars that adapt effectively to climate variation, meet consumer preferences, and address human and livestock nutritional needs. Most economically important traits in crops are controlled by multiple loci often with recessive alleles. Considering particularly Africa, this continent has several agro-

climatic zones, hence crops need to be adapted to these. Therefore, targeting specific loci using modern tools offers a precise and efficient approach. This review article aims to address how these new technologies can provide a better support to smallholder farmers.

KEYWORDS

agriculture, marker genotyping, mutation breeding, NGT, CRISPR, legislation, climate-smart cultivars

1 Addressing food crisis sustainability under the current climatic challenges

According to the FAO et al. (2019), over 820 million people globally live with hunger, which is a phenomenon rising in almost all subregions of Africa, and to a lesser extent in Latin America and Asia. Since 2014, food insecurity across the sub-Saharan Africa region, particularly South Sudan and Nigeria, has been exacerbated by human conflict as well as severe and long periods characterized by drought and sudden pests' attacks, particularly affecting pastoralists (Anderson et al., 2021). Drought is known to be one of the main factors behind the undernourishment increase in this region (FAO et al., 2019) and has been consistently identified as a primary driver of famines (Maxwell and Hailey, 2020), especially in countries such as Ethiopia, Nigeria, and South Sudan.

Yet, alongside insufficient food intake, there is a contrasting issue where both children and adults suffer from obesity, diabetes, and other food-related diseases due to the consumption of low-quality and nutritionally poor food. In fact, obesity, along with overweight, contributes to approximately four million deaths annually worldwide. The connection between obesity and food insecurity is partly driven by high food costs and the widespread reliance on inexpensive sources of fats and sugars (GBD 2015 Obesity Collaborators et al., 2017). Despite the apparent stabilization in the global hunger trend observed throughout the 2021–2022 period, the concern remains regarding the significant number of women in rural areas who continue to struggle with accessing nutritious, safe, and sufficient food (FAO, 2023). Hunger remains a persistent challenge in least-developed regions, especially across all sub-regions of Africa (FAO, 2023) because substantial gaps between potential and actual crop yields continue to hinder food security (Figure 1). Extreme weather, particularly variation on temperature and rainfall, along with limited resource availability, further contribute to disparities in crop production across the different regions. For example, rainfed upland rice systems are more sensitive to soil moisture variability when compared to irrigated paddy rice systems (Stuecker et al., 2018). Thus, selecting cultivars that meet needs and expectations while providing stable incomes is quite challenging. The impact of heat and drought on crops varies by season and geography, with evidence showing that inter-annual climate variability can significantly affect harvested yields and overall production (Roberts et al., 2009; Lizumi and Ramankutty, 2016). Only climate-smart plant breeding can provide new cultivars that are tolerant to heat and drought.

Recent findings from the Agricultural Model Intercomparison and Improvement Project (AgMIP), and Inter-Sectoral Impact Model Intercomparison Project (ISIMIP) showed that higher temperatures generally result in lower grain yields for maize, soybean, and rice. Conversely wheat yields have been found to increase due to higher CO_2 concentrations together with expanded high latitude regions (Jägermeyr et al., 2021). In the Iberian Peninsula, spring maximum temperatures contribute to significant grain yield losses in barley and rainfed wheat, but also for reducing the marketable tuber production in potatoes. However, northern regions are projected to experience increased grain yields due to early winter warming which promotes earlier growth of seedlings (Bento et al., 2021). These changes suggest that adjustments in agricultural practices, management and selection of crop species and cultivars are necessary.

2 Cause root and derived actions

The Paris Declaration endorsed the base development efforts on first-hand experience rather than simply conveying aid. It was supported by five main pillars: ownership, alignment, harmonization, result management, and accountability. However, despite these well-intentioned principles, the desire for quick results may have contributed to a lack of clarity and coordination in the development process, leading to incomplete fulfilment of providing nutritious and healthy food for all in a sustainable manner though current food systems. The impediments are many and may lie beyond 'agricultural' restrictions. Imbalances related to internal governance, either from the economic or territorial sides, do not convey the fairness of expected returns. Under a set of partnerships between the European Union (EU), the Food and Agriculture Organization of the United Nations (FAO/UN), the French Agricultural Research Centre for International Development (CIRAD), and national stakeholders, key points, including food

security, nutrition and health were gathered for inclusive and sustainable solutions to transform food systems while preserving ecosystems and landscapes (FAO, 2021).

Africa, with its 55 countries, is the continent with the highest number of countries, and the highest population growth rate. By 2100, five of the 10 top-most populous countries in the world will be in Africa, accounting for over half of the projected global population growth through 2050. Additionally, the problem of inadequate access to food in countries affected by conflict is worsened by natural disasters, economic challenges, and public health issues (UN-DESA, 2022). This means that in Africa there is a high demand for increased productivity (Figure 1), and to counteract rising food prices, annual production needs to grow substantially. In these endowers should be acknowledged multiple efforts made in water management and some plant varieties adapted to low input conditions (Nyika and Dinka, 2023; CIMMYT, 2021).

Criticism was raised when it was pointed out that major efforts to improve water and nutrient efficiency in plants have not

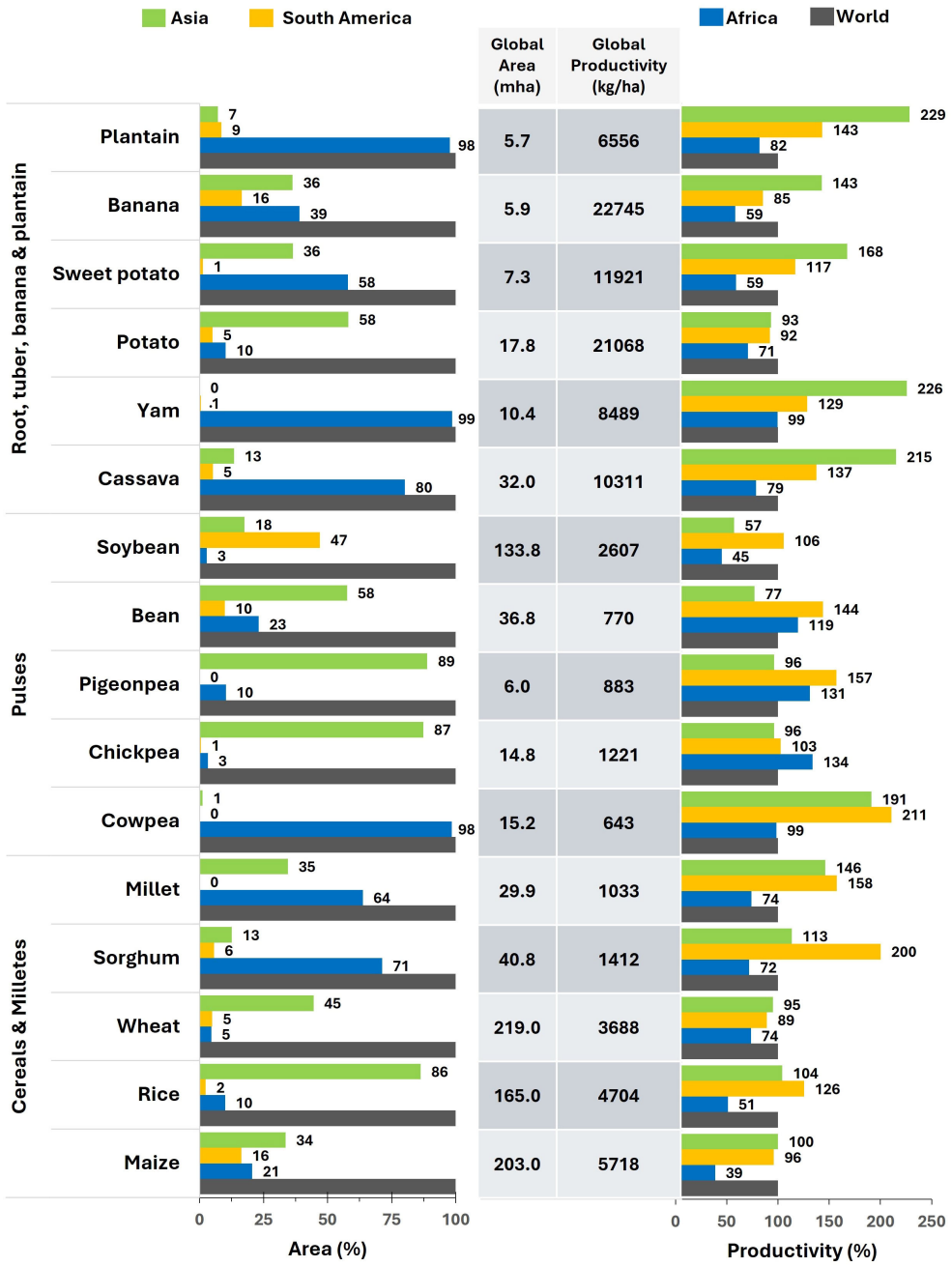


FIGURE 1
Major field crop area *vs.* productivity in Africa, Asia, South America and world. The figure illustrates that in Africa, yam, cassava, plantain, millets and sorghum are crucial crops for maintaining food security sustaining the need for more breeding investments in these crops. Pigeonpeas, chickpeas, common beans, potato, but also maize, rice and wheat are examples of crops in which productivity exceeds the cultivated area. (Source: FAODATA, 2022).

produced major results in the tropics, ignoring the fact that the advantages of crop improvement research and social protection programs are relative and dependent on the situation in a particular region (McIntire and Dobermann, 2023; Ginkel and Cherfas, 2023) (Figure 2).

In climate-smart plant breeding, the tackling of these challenges involves separate genetic selection for rainfed and irrigated conditions and for resistance to pests and pathogens. Additional actions include research efforts aimed at rejuvenating growth in crops that have reached a plateau yield improvement.

Besides the use of hybrid cultivars, the increase in grain yield in crops is known to be positively correlated with increased levels of nitrogen (N) fertilizers (Maheswari et al., 2017). Yet, despite N-fertilizers being a key determinant of yield, it also represents a significant cost for the farmers and in addition may have a negative environmental impact. An example is nitrate leaching from the field leading to soil and underground water contaminations as well as to greenhouse gas emissions. In sub-Saharan Africa, transformation in productivity can be achieved quickly through increased use of fertilizers and irrigation management. Yet, it is essential to consider that applying fertilizers without irrigation management has limited benefits, and the cost of irrigation can be a limiting factor (McIntire and Dobermann, 2023). Actions in this direction include breeding for nitrogen use efficiency (NUE: grain dry matter yield *per* unit of N availability) as well as tentative reduction of excess N-fertilizer input by management while maintaining an acceptable yield and quality. NUE is particularly important for smallholder farmers since they are known to have few resources at their disposal at any given time with which they work under very stringent financial constraints. Hence, this means that farmers cannot afford the luxury of buying large quantity of fertilizers as is always the case where one system is used for both feeding people and generating income. The trait NUE is controlled by many genes and fortunately many of these have in the recent years been genotyped and amenable for employing integrative tools for genetic enhancement of crop plants.

Optimization within breeding programs (BPs) is being led by management and optimization of the BP itself, using standardized tools and services. Yet, this may not necessarily lead to the desired improvements, as plant breeding is currently limited by the reliance on the selection of rare naturally occurring mutations in gene-regulatory regions and for traits that are controlled by multiple genes (Dwivedi et al., 2023). Rare alleles are mostly under-looked in the majority of quantitative genetic analyses since these are found in very few individuals of a population (Garcia-Oliveira et al., 2014), e.g. within less than a 2% frequency in a given population of a crop such as rice (Purugganan and Jackson, 2021).

Having a diverse set of alleles related to nitrogen uptake, assimilation, and utilization can contribute to improved NUE in plants. These alleles may also carry specific traits that make plants more adaptive to nitrogen-limited conditions. For example, in bread wheat, certain cultivars are more prone to aluminium stress because a panoply of physiological traits making plants more resistant to the damages incurred by aluminium in the soil (Garcia-Oliveira et al., 2016). In the case of N, the rare allele might encode a transporter protein with a higher affinity for nitrogen uptake from the soil, allowing the plant to acquire and utilize nitrogen more efficiently. If such rare alleles are identified and successfully incorporated in the target crop through breeding, it could result in a cultivar with improved NUE. Recently, Yoon et al.

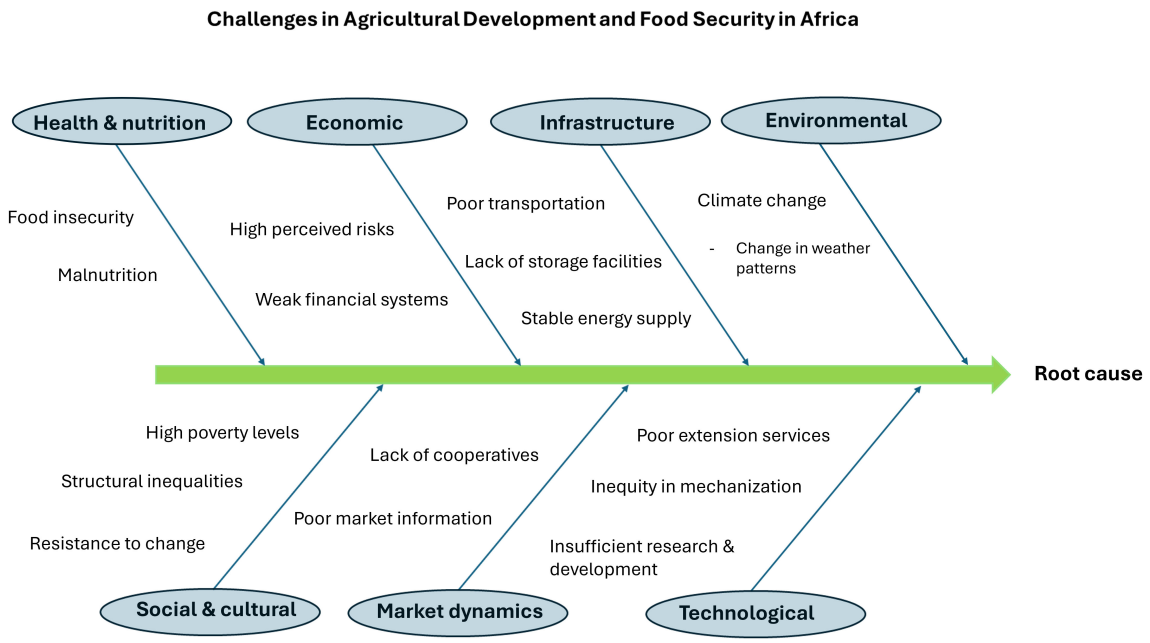


FIGURE 2
Root cause analysis of the challenges faced by the African continent in agricultural development. Additional issues stated in the backbone diagram include emphasis in quick results, confusion among the stakeholders particularly when not involving the local communities in decision making and setting priorities, as well conflicts leading to lack of trust among partners.

(2022) discovered that the nonsense-mutated *GS3* gene, represented by the *gs3* allele in the rice cultivar 'Akita 63', enhances yield production and grain size in rice, thereby improving harvest index and NUE in rice. With the newest mutational breeding tools the discovery can be transferred to other crop species swiftly.

3 Climate-smart cultivar selection

When it comes to cultivar substitution and recommendation, it is important to consider inter-variability periodicity. Adjustment to identify main parameters including genotype nature (G), environment (E), and their interaction (G×E) would improve the predictions. The main aim is to make correct forecast and decisions that would result in maximum outcomes. Besides, the ploidy level, genotypic value, and allele frequency must also be taken into consideration. When considering G×E interaction parameters related to biotic and abiotic stress tolerance, a mean performance across different trials conducted under various field management (M), effects of the trial (E), repetition (e), genotype (G), interaction of genotype k and trial i (G_kE_i), residual error, number of trials and number of repetitions per trial should be taken into consideration in the model applied. Plant breeders focus on specific traits, also known as TPTs (*T*arget *P*erformance *T*raits), when developing new cultivars with improved tolerance, for example to abiotic stresses. Defining these traits, which form the basis for the ideotype concept (Donald, 1968), is crucial for effectively selecting and breeding plants that can withstand challenging environmental conditions. With reference to African climate and conditions, the G×E effects may be hugely underestimated for operation efficacy, along with a lack of resources. More precisely, four major agro-climatic zones, can be considered, and each characterized by distinct climate, soil, and vegetation types, namely the a) Arid and Semi-Arid Zone with low rainfall and high temperature (e.g. Sudano-Sahelian zone), the b) Sub-Humid Zone where the rainfall frequency is more reliable (e.g. coastal areas), the c) Humid zone characterized by high rainfall and supporting lush vegetation and diverse crops (e.g. the Central and southern parts of West Africa), and finally the d) Tropical high-altitude zone, typically comprised of highland areas characterized by cooler temperatures and distinct growing seasons (e.g. East African highlands of Ethiopia and Kenya as well the Rift valley of Tanzania). Studying these interactions may be very demanding, but principal component analysis (PCA) may help in this multifactorial-dependent process. Taken this continent as an example breeder's face challenges specific to each climate zone where water scarcity, pests and disease pressure, and soil problems vary within these zones. Developing climate-smart cultivars, such as drought-resistant varieties for arid and semi-arid zones, involves more than just adaptation to a changing environment. It must first prioritize customization to the consumer's needs and preferences. These market choices drive producers to use and harvest cultivars fitted to the unfavourable environment due to exposure to erratic edaphoclimatic conditions in terms of frequency and intensity (Table 1). For cultivar selection a significant genetic variation is needed, and where germplasm diversity is met by the high phenotypic variability provided by wild relatives, old cultivars, and landraces into elite lines with a known and stable yield production background (Garcia-Oliveira et al., 2009). Because

TABLE 1 Predicted impact on global agricultural potential based on the projected trends in climatic change scenarios and its management through mitigation/adaptation practices.

Factor	Event	Potential impact	Likelihood	Mitigation/adaptation Practices
Temperature	Cold periods becoming warmer and shorter; over most land areas, days and nights becoming hotter.	Positive impact on crop yields in cooler environments particularly northern hemisphere countries, but it is likely to be either offset or severe reduction in warmer environments (tropical and sub-tropical), increased outbreaks of new insect pests and pathogens resulting in potential impacts on crop production; change in the life cycle of bees, and as consequence, in pollination and plant fertility.	Virtually certain	Crop management practices: Ecosystem-based integrated nutrients (soil) and pests (diseases, insect-pest and weed) management approaches such as conservation agriculture, alterations in cropping patterns and rotations, crop diversification, mulch cropping, cover cropping, organic agriculture, irrigation management and land fragmentation among others. Crop improvement practices: ✓ increased access to high-quality seeds/planting materials of adapted varieties, ✓ closing of yield gap in developing and least developed countries through rapid integration of new genomic technologies and ✓ development of improved site-specific crop varieties
Precipitation	Heavy precipitation events increasing in frequency over most areas.	Damage to crops; soil erosion; inability to cultivate land owing to waterlogging of soils.	very likely	
	Drought-affected area increases.	Land degradation and soil erosion; lower yields from crop damage and failure; loss of arable land.	Likely	
Air	Intense tropical cyclone activity increases.	Damage to crops; change in the normal life cycle of pests and diseases, changes in the spatial distribution of weeds such as *Striga hermonthica* in cereals.	Likely	
	CO_2-induced warming resulted in rising of sea levels (excludes tsunamis).	Salinization of ground water in coastal area, estuaries, and freshwater systems; loss of arable land due to inundation of low-lying area.	Likely	

climatic conditions evolve, cultivar replacement is thought to be a more frequent need when the primary focus is on supply-driven concerns. More focus should be put on adding locations (multi-site trials) as opposed to years (multi-year trials) (van Etten et al., 2019), through farmer-participatory on-farm trials. The advantages to this approach include that location variables tend to be more predictable than year variables, and one could target locations with different variable profiles to get well-adapted cultivars. Additionally, it attracts a greater number of participants and affords grower cooperators a meaningful presence in the discussions or decision-making processes. Yet, decisions from the breeding side must take into consideration their resource allocation of what is feasible in terms of number of years, number of sites, and how to choose sites to maximize adaption/climate-smartness.

4 Toolbox for inclusion of climate-smart traits

Despite considerable advancement during the last decades, in terms of genomic tools and services, and marker utilization *per se*, still there is a low level of integration of molecular markers, in plant breeding programs. Several reasons could explain this, particularly in Africa, namely,

i. Establishment of genotyping facilities, accessible to all breeding programs is costly
ii. Lack of physical infrastructure such as uninterrupted power supply, centre for data storage, among others.
iii. Issues related to availability of qualified, cost-effective human resources
iv. High cost of genotyping due to underutilization of genotyping facilities
v. Inability to deliver high-quality genotyping data within a short period to meet breeding decision timelines
vi. Lack of comprehensive understanding of the potential of molecular markers in breeding programs
vii. Low or no funding for crop improvement from governments

To foster the integration of genomic data tools in breeding programs there must be a change in mindsets. This will require us to provide genotypic information consistently at a high-standard pro prompt and comprehensive interpretation, to assist in the implementation of advanced techniques such as marker-aided selection and genomic prediction of breeding values for further use in selection. It is very important to implement rigorous quality assurance processes that use Quality Control (QC) molecular markers for data integrity be maintained. Identifying more efficient methods of genotyping which will help in reducing the costs related with data generation and analysis. This will ensure that genotypic data is available within the breeding program workflow in time for decision-making. Additionally, it is important not to forget about embracing technological advancements in such areas as the utilization of drone-based imaging sensors and genetic algorithms based on artificial intelligence, for high-performance phenotyping since these breakthroughs can remarkably improve efficiency and precision of BPs. Concerning the implementation of genomic technologies, several questions should be posed before starting using genomic technology, including how the efficient and effective application of genomic technology can be leveraged to support BPs. Hence, there are three questions (3Qs) to address in planning a breeding program (Figure 3).

Advanced technologies such as Kompetitive Allele-Specific PCR (KASP) markers, high-throughput genotyping, double haploid (DH) technology, mutational signature markers, and clustered regularly interspaced palindromic repeat (CRISPR)/CRISPR-associated endonuclease 9 (Cas9) technology must serve the purpose and be applicable in a breeding program. However, certain conditions must be met for this to occur. These prerequisites involve the availability of reliable markers linked to specific traits, the access to mutational signature markers, as well the presence of advanced laboratory infrastructure and equipment conducive to the efficient utilization of CRISPR technology.

4.1 Publicly available marker panel platforms

With the advances in technology, particularly deep sequencing, the lowering of cost allowed the validation of other molecular panels to the crop community (Table 2). Several publicly available molecular marker platforms for plants include Gramene, Plant Markers Database (PMDB), and LegumeSSRdb. Gramene offers the access to SNP and SSR markers (https://www.gramene.org/; accessed November 3, 2024), whereas PMDB provides pathway-based markers for 82 plant genomes, including maize, rice, and sorghum (http://ppmdb.easyomics.org/; Mokhtar et al., 2021). In its turn, LegumeSSRdb is a comprehensive resource for SSRs in thirteen legume species (http://bioinfo.usu.edu/legumeSSRdb/; Duhan and Kaundal, 2021). Another straightforward example is the One CGIAR-CIMMYT marker platform. Initiated by the CGIAR Breeding Platform (https://excellenceinbreeding.org/toolbox/tools/kasp-low-density-genotyping-platform). This platform toolbox gathers several sets of useful markers developed for different traits and quality control (QC) in numerous crops to assess and ensure the quality and purity of breeding materials through sharing arrangements with dozens of scientists. Further, to validate the trait-specific KASP-based markers, the platform also includes the names of validated plant mid-density panels using Diversity Array Technology tag (DArTtag). Besides the above mentioned molecular marker types, there was a subsequent surge in the availability of similar platforms including the *Bean Improvement Cooperative* (http://www.bic.uprm.edu/?page_id=91) for common beans, the OZ Sorghum (https://aussorgm.org.au/sorghum-projects/) for description of quantitative trait loci (QTL) and providing resources for further marker sequence design, as well

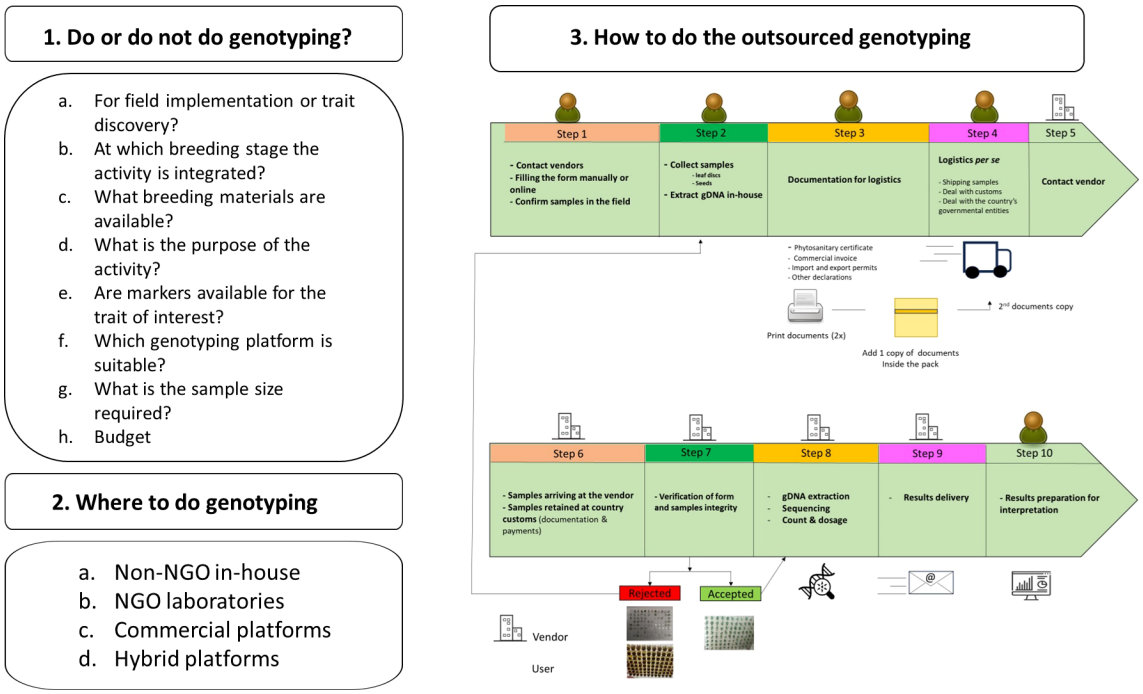


FIGURE 3
Example of workflow assessment for genotyping implementation in a breeding program (BP) as a part of a routine process in the breeding operations. Starting by defining the ideotype including trait target values, followed by accessing if genotyping is feasible, followed to understand where to perform the activity and the steps involved at the general level when outsourcing of genotyping is involved. Logistics and phytosanitary regulations are two major bottlenecks, with each country having different regulations for the same process.

Primer3 (https://junli.netlify.app/apps/design-primers-with-primer3/) for free KASP markers design. A more recent, but equally interesting platform, is the Flex-Seq panels from Biosearch™ Technologies (https://www.biosearchtech.com/flex-seq-industry-standard-pre-designed-panels-8262) which allows different marker densities, and after made up of a standard panel for a crop any number of markers (sub-panels) can be used according to the needs of the breeding programs. Examples of these panels include maize, potato and soybean.

4.2 Process to incorporate molecular markers into BP workflow

The basic steps of crop improvement include parental selection and crossing where there is a need for continuous genetic variation assessment and progeny selection. Without this yearly assessment, breeding programs cannot assure the suitability of their materials for trait improvement. Steps to integrate molecular markers include marker discovery, validation, and development (Figure 4) and many

TABLE 2 Comparative overview of the most common PCR based DNA markers in plants.

	RAPD	AFLP	SSR	SNP
Inheritance	Dominant	Dominant/Co-dominant	Co-dominant	Co-dominant
Genomic abundance	High	High	Medium to high	Very High
DNA quantity required	Low	Medium	Low	Low
DNA quality	Lower than AFLP	Higher than SSR	Lower than SNP	High
No. of polymorphic loci	1.5-50	20-100	1.0-3.0	Thousands
Reproducibility	Low	High	High	High
Accuracy	Very Low	Medium	High	Very High
Development effort	Very Low	Low	High	High
Development cost	Low	Medium	High	Low
Running cost	Low	Medium	Low	Low
Automation	High	Low	High	High

RAPD, Random Amplification of Polymorphic DNA; AFLP, Amplified Fragment Length Polymorphism; SSR, microsatellite or Simple Sequence Repeat; SNP, Single Nucleotide Polymorphism. The number of polymorphic loci refers to the total count of loci that are polymorphic.

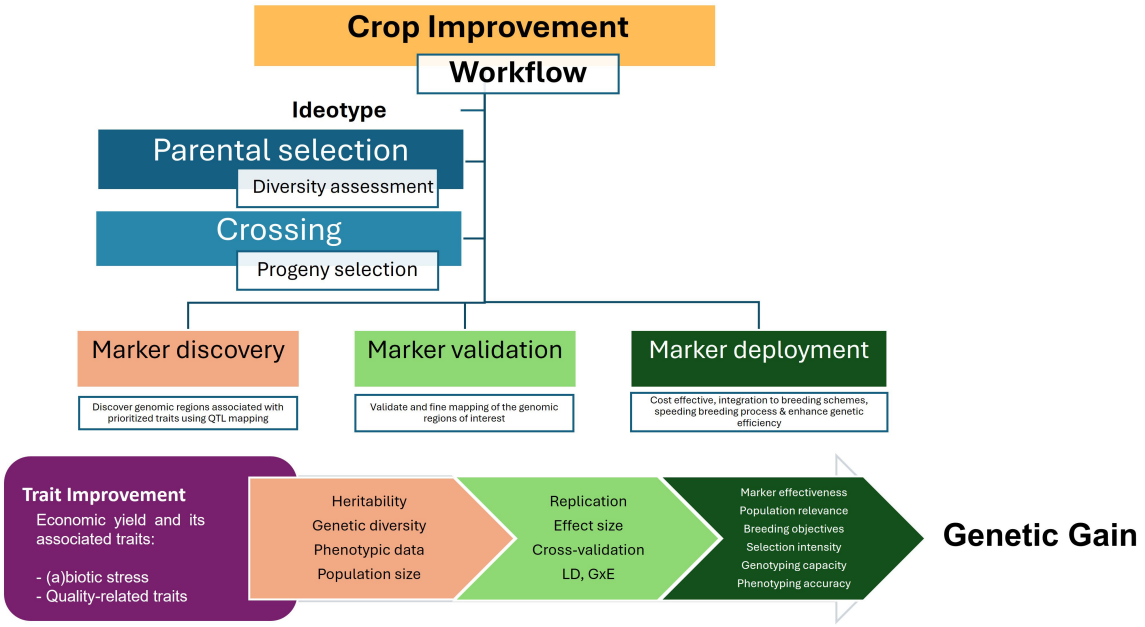


FIGURE 4
Process to incorporate molecular markers into BP workflow. The discovery, validation and deployment of molecular markers is applicable across crop plant species.

are the successful examples on how to apply them in BPs through diverse partnerships (Garcia-Oliveira et al., 2021, 2022, 2024).

In marker discovery, the genomic regions affecting the trait of interest for improvement are discovered through QTL mapping, including QTL identification and genome-associated mapping (GWAS), or by using reverse genetics. Once the relevant loci are revealed, they need to be validated in a wider range of targeted populations and optimized for technical efficiency. The marker validation process may involve the fine mapping of relevant loci. Finally, the validated markers are deployed in a cost-effective manner and integrated directly into the targeted breeding programs to enhance efficiency of genetic gain for specific trait.

The University of Minnesota's (UoM, St. Paul, Minnesota, USA) Wheat Improvement Program uses markers in three main ways: 1) Parental genotyping to facilitate crossing decisions, 2) Enrichment of 3-way crosses for favourable marker alleles prior to generation advancement, and 3) Marker-assisted line purification during advanced breeding stages. Crossing parents are screened with KASP markers for approximately 50 genes influencing agronomic, end-use quality and disease-resistance related traits. Since parental genotypes are known, 3-way cross progeny can be screened and enriched for favourable alleles of segregating markers. In marker-assisted line purification, following generation advancement, lines are planted in plots of four head-rows, observed phenotypically, and screened with around a dozen KASP markers to select the most uniform and typical row to carry forward as the line. Markers are also used to screen bulks of large-scale seed increases, and if segregating markers are found, individual rows within the plot are screened to facilitate the grouping of impurities. In the case of marker-assisted line purification, special care is taken to precisely follow best experimental practices due to the importance of accurately identifying segregating rows. DNA is carefully normalized, and control genotypes for each allele, as well as negative controls, are included in each KASP experiment to facilitate accurate clustering (*personal communication Dr. Emily Conley, UoM, Minnesota, USA*).

In the breeding program of the International Potato Center (CIP, Lima, Perú), for the vegetatively propagated potato, DArTag markers (Endelman et al., 2024) are being used for genomic selection (GS), diversity analysis, and selection based on trait markers. KASP markers are used for traits with no available SNP markers in the panel array, as well to monitor the genetic identity of plants used in crossing plants (Kante et al., 2021). In Yam at the International Institute of Tropical Agriculture (IITA, Ibadan, Nigeria) KASP SNP markers are being used to distinguish different yam species (*Diascorea alata*, *D. rotundata*, *D. praehensilis*, *D. cayenensis* and *D. abyssinica*, and also tagged markers with the economically important traits such as plant vigour, sex distinguishing, flowering intensity, yam mosaic virus, among others (Agre et al., 2023). For marker discovery, other platforms with higher marker density can be of use.

5 Identifying trait-associated haplotypes for improved breeding

5.1 KASP-haplotype and epiallele underlying quantitative traits

In breeding, trait-associated haplotype identification is valuable in the context of genetic selection through the manipulation of

population allele diversity. The inclusion of haplotype information in genomic selection models can enhance the accuracy of predicting an individual's breeding value (Sallam et al., 2020). Haplotypes not only capture more extensive genetic variation but also provide a finer resolution of genomic variation compared to individual SNPs alone. KASP-haplotype genotyping allows the detection of multiple SNPs within a haplotype, thereby improving linkage disequilibrium between causal loci and marker haplotypes and leading to more precise estimates of marker effects for genomic selection.

However, to date, from QTL mapping, thousands of loci have been described for numerous traits, many of which control only a small proportion of the trait variance. Not only is the true proportion of phenotypic variance these KASP loci explain unknown but also these haplotypes of interest change from breeding cycle to breeding cycle due to recombination and introduction of new germplasm. The use of KASP-haplotypes can capture rare genetic variants that might not be detected or accurately imputed using individual SNP markers. Additionally, in the majority simply inherited traits, haplotypes are easy to find and quantify their contribution to the trait phenotype. Yet, the genetic complexity of the target trait, the genetic diversity of the population, the trait architecture, and the availability of genomic data, all factor into the ease of identifying and quantifying haplotypes of interest. A similar 'in-house made' method to KASP methodology is the "Amplifluor-like" method, and competitive against KASP methodology. A primary advantage of the Amplifluor-like method lies in its unparalleled flexibility, allowing for the modification or rearrangement of all its elements. This includes the ability to alter the structure and design of allele-specific primers and universal probes, as well as the amplification program, including parameters such as temperature and duration for each step (Khassanova et al., 2023). With the availability of knock-off mastermixes, that can be cheaper, and the primers ordered from whoever provides the cheapest oligo synthesis service in a particular region, one that has already an equipped and expert laboratory, such one at the UoM can design allele-specific assays (Allele 1, Allele 2 and reverse primer), add the hex tail on one primer and a fam tail on the other, and use PACE mastermix, which is commercially available but for about half the price of KASP (https://3crbio.com/). However, in breeding programs where laboratory facilities and consumables purchase are not competitive, using of shared services available is ideal.

Another motive to consider is the use of epialleles, as these can play a great role in the regulation and contribution to observed variation of quantitative traits (Gahlaut et al., 2020). Epigenetic modifications can influence gene expression levels, developmental processes, and responses to environmental cues, thereby impacting complex traits. Understanding the epigenetic regulation not only provides valuable insights into potential avenues for breeding, but also, hypothetically, and in the context of genomic selection, genomic selection models can potentially improve the accuracy of predicting breeding values, and therefore, enhancing selection efficiency, particularly at the level of environmental responsiveness and phenotypic plasticity (Figure 5). This approach may require a different type of marker platform entirely to get genome-wide coverage of SNPs in methylated as well as unmethylated regions. Current technologies could prove challenging, as most current high-throughput platforms tend to target unmethylated (presumably gene-rich) regions. Nevertheless, the incorporation of epialleles into genomic selection is still evolving, with the need to research effective methods for integrating these elements while understanding the complex

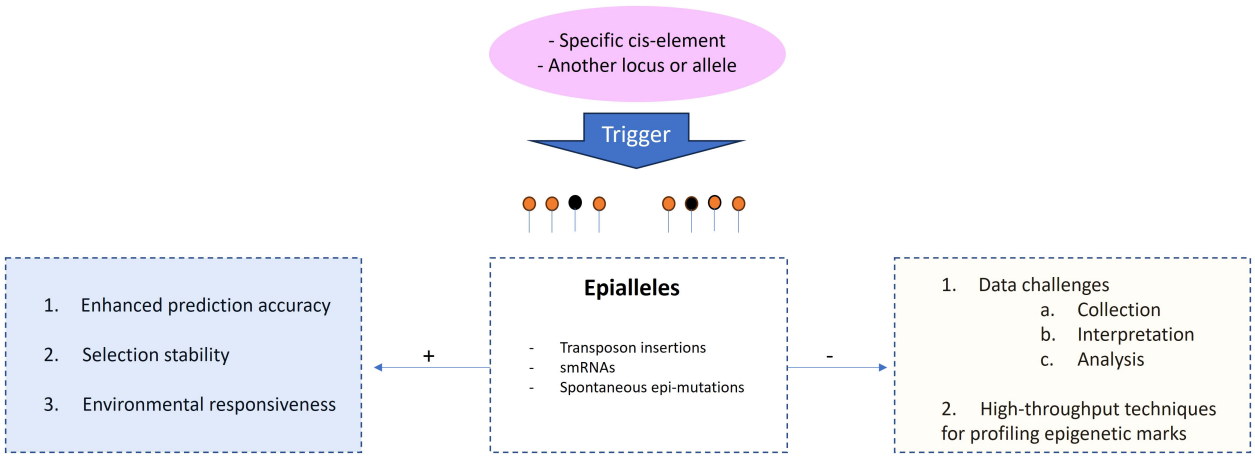


FIGURE 5
Relevance of epialleles to genomic selection. 1. Including epigenetic information, such as DNA methylation patterns associated with quantitative traits, can provide additional information beyond the DNA sequence itself, 2. Epigenetic modifications can contribute to phenotypic stability across generations by providing an additional layer of heritable variation that is less affected by short-term environmental fluctuations. When considering epialleles associated with stable phenotypes, genomic selection can facilitate the breeding of more resilient and adaptable individuals. 3. Epigenetic modifications are sensitive to environmental cues and can mediate phenotypic plasticity. Incorporating information about epigenetic marks that respond to specific environmental conditions, genomic selection models can account for genotype-environment interactions and improve prediction accuracy in varying environments. Despite continuous technological advance, careful experimental design and computational analyses are still required to extract meaningful information for selection purposes, as in the case of DNA methylation sequencing. The signals (+) and (-) refers to advantages and drawbacks, respectively.

interactions between epigenetics, quantitative traits, and breeding outcomes. Therefore, it is an exciting avenue for further improvement in breeding strategies.

5.2 Challenges when utilizing KASP markers in polyploid crops

An effective KASP SNP result is achieved when the marker demonstrates clear differentiation between homozygous and heterozygous alleles in the genotyping output. However, not all the results are straightforward, often leading to complex outcomes where alleles cannot be assigned to specific allele groups. Also, in polyploid crops, information on allele dosage might not always be available. In such instances, a process of SNP recalling becomes necessary, which in some cases is a challenging task that demands technical expertise and is often carried out manually, consuming significant amounts of time and the need for knowledge expertise. Beyond assessing homozygosity/heterozygosity, KASP genotyping data can be used to predict target observable traits (phenotypes) associated with specific genetic variations and based on known associations between certain alleles or SNP markers and phenotypic traits of interest, established through genetic studies and BPs.

Another matter of importance is marker data imputation, either in diploid or polyploid crops, to fill in the missing or incomplete genetic information in a dataset. It involves predicting or estimating the allelic variants (genotypes) of individuals at specific physical positions in their DNA sequence where data is missing. Genotypic data can be incomplete for various reasons, such as genotyping errors, low sequencing coverage, or missing samples. This practice is particularly valuable to enhance statistical power and to improve the ability to identify genetic associations with traits. Common imputation tools and software packages used in genomics research include IMPUTE (Marchini et al., 2007), BEAGLE (Browning and Browning, 2009), LD-kNNi (Money et al., 2015) and Minimac (Howie et al., 2012). It is not uncommon to use the imputation replacing missing data with the average dosage for a specific site and population, noting that the accuracy of imputed genotypes depends on the quality and size of the reference panel and the specific imputation algorithm used.

It is important to acknowledge the challenges associated with obtaining accurate dosage information when employing KASP markers for QC purposes. These challenges encompass the optimization of assay conditions, including primer concentrations, annealing temperatures, and PCR conditions. Additionally, primer design and specificity must be carefully considered to prevent cross-reactivity with other genomic regions, which could potentially yield false results. The quality and quantity of DNA are crucial factors, as contaminants, degradation, or insufficient DNA concentrations may compromise the reliability of dosage information. Sample variability poses another challenge, emphasizing the necessity to confirm that the chosen marker is suitable for the specific genetic background under investigation. The absence of reliable reference materials or standards for use as control samples can further complicate dosage determination. The presence of PCR inhibitors in the DNA sample, if not addressed, can impact the efficiency of the reaction. Moreover, the calibration of equipment is paramount, as inaccuracies in temperature control and fluorescence detection can lead to unreliable results. The analysis of KASP data requires robust software capable of accurately calling genotypes and dosage information. Finally, incorporating replicates and positive/negative controls in experimental setups is essential for assessing assay reproducibility and detecting potential issues.

On the quality side, and to enhance reliability, the use of control genotypes is recommended. Yet, achieving marker stability is vital, given that the same marker may yield different calls across various plates, even when considering the same sample repetition. This could be due to poor DNA quality, or in the case of polyploids, failure to normalize DNA samples to a consistent concentration. It could also be due to a low level of true residual variation at some loci following generation advancement and line purification. Consistent marker performance over time is also a crucial characteristic of high-quality KASP markers. This necessitates the identification and selection of markers that amplify well and cluster cleanly, especially in the context of defining a limited set of QC markers, as demonstrated in the case of soybeans (Chander et al., 2021).

6 Precision Plant Breeding with application of genotyping technologies

6.1 Mutation breeding in genomic era

DNA mutations are the basis of all heritable variation in organisms, providing the raw material for natural selection and evolution. However, such naturally stable mutations are relatively rare in eukaryotes compared with prokaryotes, partly due to the presence of complex DNA repair mechanisms in the former. In plants, mutation rates vary across different regions of the genome, with lower frequencies observed in areas under strong functional constraints (Quiroz et al., 2023). Specifically, the mutation rates are reduced by 50% within gene bodies and by two-thirds within essential genes, as compared to non-essential genes (Monroe et al., 2022). This reduction is likely due to the elevated activity of targeted DNA repair mechanisms in these regions. A deeper understanding of mechanisms that promote specific DNA repair could help achieve critical goals in crop improvement programs through targeted mutagenesis. The application of mutation techniques by plant breeders has a long history, dating back to 1920's when X-ray induced mutagenesis was first realized on fruit fly (*Drosophila melanogaster*) and cereal species (maize and barley) by Muller (1928) and Stadler (1928a, b), respectively. Muller's groundbreaking research revealed that genes typically exhibit a mutation rate ranging from 10^{-5} to 10^{-6} per locus per generation. The advent of radiation-induced mutagenesis allowed scientists to achieve higher mutation rates than those occurring spontaneously leading its application in inducing novel genetic variability through radiation in plants. Over the past century, various physical agents like X- and gamma-rays, UV light radiations, as well as particle

radiations including alpha- and beta-particles, neutrons, and protons, along with chemical substances such as alkylating agents (e.g., EMS - ethyl methanesulphonate, NEU - nitrosoethyl urea, MNU - N-methyl N-nitrosourea), colchicine, EI (ethylene imine), and sodium azide (NaZ), among others, have been extensively employed. These tools have not only been instrumental for studying gene function and DNA repair mechanisms but have also been pivotal in generating novel cultivars in various crops (Penna and Jain, 2023). Globally over 3400 cultivars derived from approximately 228 plant species have been developed in 75 countries to date, utilizing a range of mutagens (http://mvd.iaea.org accessed in January 2024). Among these, radiation-induced mutagenesis, particularly with gamma rays, remains one of the most prevalent methods.

More recently, the application of accelerated particles including heavy ions (C-, H-, and Ar- ions) or protons, has gained importance for developing novel cultivars with desired traits. These particles induce mutagenesis with high biological effectiveness at low radiation doses minimizing impacts on other phenotypes (Tanaka et al., 2010; Abe et al., 2015). The precision and specificity of accelerated particles along with their energy, reduce off-target mutagenesis making them a powerful tool in plant breeding. Similarly, space breeding conducted in environments beyond Earth's atmosphere offers another promising avenue for creating novel mutants. This approach takes advantage of two unique factors: exposure to cosmic rays and microgravity (Liang et al., 2009; Zhao et al., 2010; Liu et al., 2021a).

In any of mutagenesis methods, it is crucial to consider the direct inheritance of the DNA damage caused by the mutagen as well as the persistence of the genetically induced alterations. This phenomenon, known as mutagen-induced genomic instability can occur in the progenies that have undergone irradiation (Ma et al., 2021). Compared to accelerated particles, X-rays induce a greater extent of DNA damage and offer fewer opportunities for repair by DNA polymerases. When DNA polymerases are less efficient at correcting errors, the mutation rate increases, particularly in the case of double-strand break (DSB) mutations. This can lead to chromosomal re-arrangements and larger deletions (Pastwa et al., 2003; Tsuruoka et al., 2008; Hirano et al., 2015). High-energy particles have been used to induce qualitative traits (sterility in fruit trees and flower appearance such as colour and shape in flowers), and quantitative agronomic traits in cereals and vegetable crops (Abe et al., 2012; Kato et al., 2016; Yamatani et al., 2018; Nishio et al., 2024). The new traits obtained through these methods are useful for gene function mining, gene mapping and the development of elite alleles. Besides seed-propagated crops, *in vitro* mutation breeding in vegetatively propagated crops has also resulted in lines exhibiting several desirable traits (Suprasanna et al., 2015; Maurya et al., 2022; Zhang et al., 2024). However, there is still much to be improve in this field of mutagenesis, including optimising physical radiation parameters such as dose and high linear energy transfer (LET), understanding the impact of different irradiated particles to decrease mutation randomness as well the integrating genomic and phenomics tools to enhance mutation breeding efficiency (Nielen et al., 2018).

6.2 Next-Generation sequencing platforms

The establishment of next-generation sequencing (NGS) platforms in the recent advances of technology has enhanced our capacity to sequence and map crop genomes, identify causal mutations and study gene regulation. However, it should be noted that the majority of these genomes are complex and characterized by a relatively high proportion of repetitive sequences and transposons. For larger genomes, the short sequence reads (<700bp) generated by second-generation sequencing platforms like Illumina, Roche, Solid, and others may not be ideal. The second-generation sequencing can be extremely useful but requires methods to reduce complexity and target informative regions. For example, Illumina NovaSeq platform can be used for routine genomic prediction in a breeding workflow, as it is low cost, despite requiring specialized knowledge to prepare libraries, access to core facilities with the sequencing instruments, and bioinformatics expertise to process the data. In cases where this is not possible long-read sequencing such as Oxford Nanopore and PacBio, also understood as the third-generation sequencing generation platforms, need to be used. These long-read sequencing platforms offer the capability to sequence longer fragments, enabling more comprehensive coverage of large genomes, especially in highly variable genomic regions which may facilitate the identification of epigenetic marks such as DNA methylation and gene expression analysis. However, these third-generation platforms are more expensive and generally less available than those of the second-generation. These are useful for discovery-type research, whereas second-generation would be more useful for routine breeding applications. For the detection of rare point mutations in plant genomes the use of NGS still presents challenges. As an alternative, the Simple, Multiplexed, PCR-based bar-coding of DNA (SiMsen-seq) system is an opportunity for selective mutation detection using sequencing, as it detects variants at or below 0.1% frequency with low DNA input (Ståhlberg et al., 2017).

Generally, mutation breeding operates outside of the typical regulatory controls. To make breeding more efficient, molecular markers can be used in selecting materials for specific traits of interest, which are then further identified in the field. Afterwards, characterization of these mutants can be carried out using advanced NGS pipelines. More importantly, it should be noted that genome editing technologies can play a significant role in trait pyramiding by enabling the introduction of multiple site mutations for different desired plant characteristics (Figure 6).

6.3 Genomic tools for utilization of microRNAs and their native targets

MicroRNAs (miRNAs) are a type of short (19-24 nucleotides in length) non-coding RNA molecules which are present within the inter and intragenic regions of genomic DNA. Since the discovery of first miRNA *lin-4* in the nematode (*Caenorhabditis elegans*) about 30 years ago, much evidence has shown that miRNAs are ubiquitously present in eukaryotic genomes that regulate gene expression, mainly through sequence-specific cleavage and/or

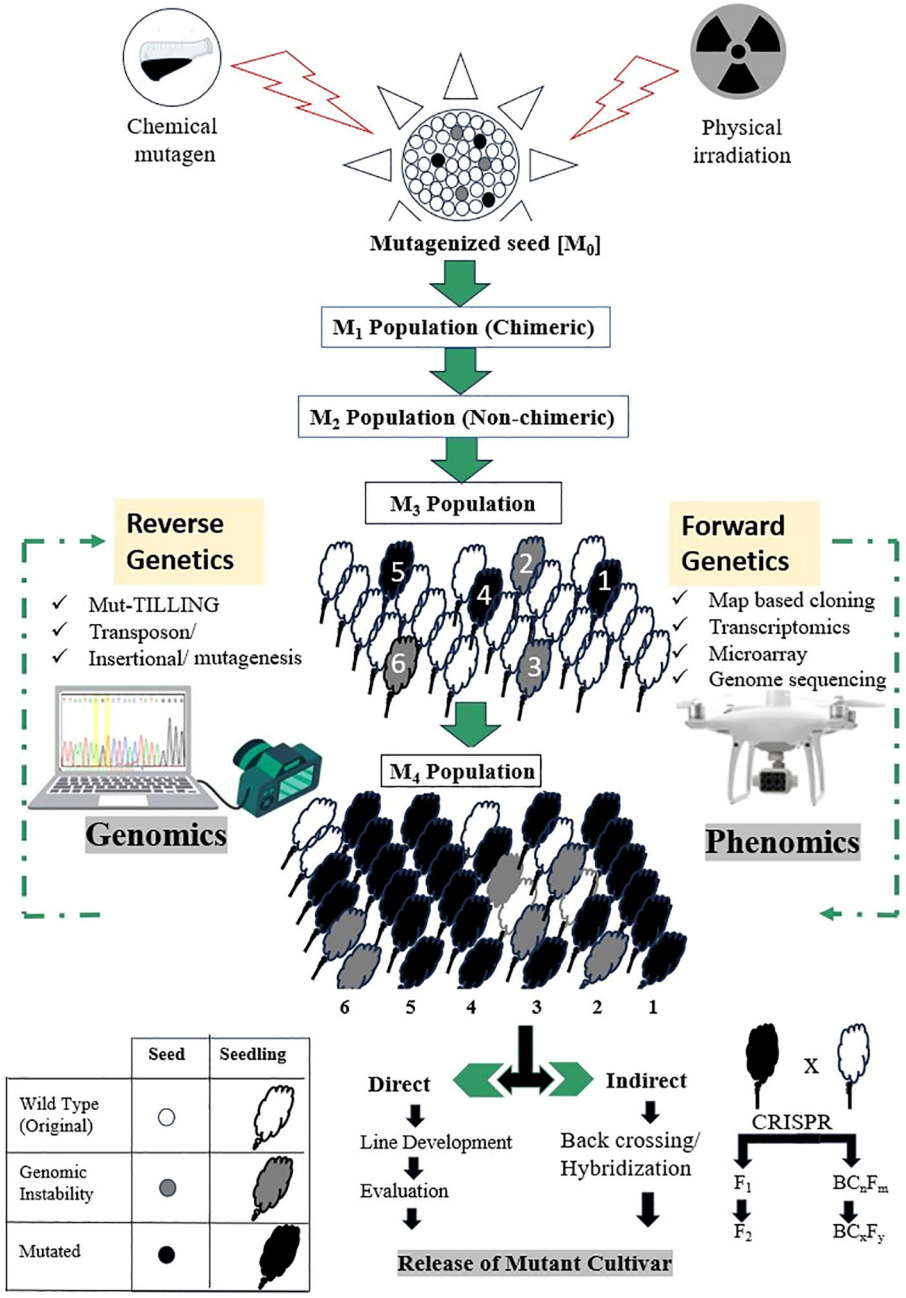


FIGURE 6
Use of high throughput genotyping and phenotyping for the development of desirable cultivars in plant mutation breeding. In forward genetics, after applying mutagenesis (such as radiation) and selecting M1 seeds, performance of precise phenotyping is done at the M2 generation. By the time the M3 generation is formed, a more refined phenotypic evaluations can be done, as the M3 plants are typically more stable and can provide clearer insights into the genetic basis of the traits of interest. This approach helps in identifying and confirming the relationship between specific mutations and the observed phenotypes. In reverse genetics, the focus is more on specific genes and their functions rather than broadly phenotyping entire populations like in forward genetics, therefore, CRISPR and gene editing can be used initially to create mutations in specific genes or may induce mutations through radiation. This is followed by phenotypic evaluation for change, starting at M2, and where the effects of mutations can be observed. By M3 generation, stable mutants can be characterised to confirm the function of targeted genes based on the phenotypes observed.

translation inhibition of the target mRNAs during or after transcription (Lee et al., 1993; Wightman et al., 1993; Voinnet, 2009; Sanei and Chen, 2015). As a result, miRNAs are emerging as master modulators of gene expression that could be utilized as next generation genomic tools for improvement of traits of interest in agronomically/economically important plants (Djami-Tchatchou et al., 2017; Tang and Chu, 2017; Raza et al., 2023). Most of functionally validated miRNA families are not only evolutionarily conserved among plant species, but also tend to have conserved targets which make more convenient of miRNAs for application in

trait improvement in crop plants (Tang and Chu, 2017). Various types of miRNAs, including conserved miRNAs that facilitate comparative studies across species, tissue-specific miRNAs that inform on developmental processes or stress responses, and stress-responsive miRNAs that are upregulated or downregulated in response to stresses, have the ability to enhance their utility in breeding stress-resistant varieties. To date, the utilization of natural genetic variation in molecular marker-aided breeding mainly relied on either random DNA markers or so-called functional markers that derived from protein coding genes.

The continuous progress in DNA marker technology replaced previous PCR-based genotyping methods. High-throughput sequence-based single-nucleotide polymorphism (SNP) markers emerged as an attractive option because of low genotyping error rate, and amenability to automation, thereby resulting in a drastic reduction in cost per data point (Chander et al., 2021). SNPs in miRNA (MIR) genes and their target genes are widespread and can influence miRNA biogenesis and function by altering transcriptional patterns or miRNA-target interactions (Todesco et al., 2012; Liu et al., 2014). In rice, nucleotide polymorphism (GG/AA) in the terminal loop region of the osa-miR2923a precursor could be differentiated from japonica/indica cultivars which are also found to be significantly associated with grain length characteristic (Wang et al., 2013). Similarly, the naturally occurring nucleotide variation in OsmiR156h, exhibited reduced plant height, enhanced lodging resistance, increased tiller numbers per plant, and resulting in an increased grain yield (Zhao et al., 2015). Few recent examples, and exemplifying their outcome importance, include the miRNAs OsmiR168 and OSmiR395, targeting the Ago1 and ATP sulfurylase (*OsAPS1*) genes which enhanced the resistance and broad-spectrum resistance to rice blast fungus (*Magnaporthe oryzae* L.) and both pathovars *Xanthomonas oryzae* pv.oryzae (Xoo) and *X. oryzae* pv. oryzicola (Xoc), causing bacterial blight and leaf streak diseases in rice, respectively (Wang et al., 2021; Yang et al., 2022). Similarly, in alfalfa miR156 target combined stress modulators. The miR156 not only was found to module flooding tolerance by regulating physiological processes and SnRK1 gene expression (Feyissa et al., 2021) but also to influence heat, cold and drought tolerances by downregulating the SPL gene (Arshad et al., 2017a, b; Arshad and Hannoufa, 2022; Yadav et al., 2024). Therefore, the integration of NGS information together with various bioinformatics tools and transcriptomic studies may aid in the identification of the functional role of natural or induced variations at miRNAs and their target loci that could be utilized to develop molecular markers. For instance, in the case where the target gene has a desirable effect on the trait of interest but MIR serving as a negative regulator, selection of natural or induced miRNA-resistant target gene mutant strategy can be used. Contrarily, if a candidate MIR gene serve as a positive regulator and its native target gene has an undesirable effect on the trait of interest, the natural or induced MIR variant, such as the suggested isoMIR, can be utilized (Morin et al., 2008). It should be noted that compared to canonical miRNAs (https://www.mirbase.org/), their isoMIRs have often shown differential expressions, higher binding capacity, and efficiency in target cleavage in *Arabidopsis thaliana* (Jeong et al., 2013; Ahmed et al., 2014).

6.4 Genome editing tools for crop improvement

Using high-throughput sequencing (HTS), hundreds of miRNAs have been identified and annotated in different plants species (https://www.mirbase.org/), but their biological function still unknown. Like the study of protein-coding genes, following the identification of miRNAs-targets, their functional characterization is necessary by the creation of gain-of-function or loss-of-function analyses. To perform gain-of-function studies, miRNA genes or miRNA-resistant target genes could be overexpressed using various techniques such as traditional constitutive promoter, full-length cDNA of miRNA genes, or the two-hit artificial miRNA vector system, which is highly adaptable to plant transformation techniques (Teotia et al., 2020). As a result, most of miRNA studies have been conducted on Arabidopsis due to the ease of the floral-dip transformation method (Zhang and Unver, 2018).

While RNA interference (RNAi) and site-directed genome engineering techniques such as zinc finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs) approaches were successfully employed for loss-of-function studies (Curtin et al., 2011; Li et al., 2012; Ainley et al., 2013; Liang et al., 2014). Compared to traditional genetic mutants of protein-coding genes, miRNA mutants are less effective due to the small size of miRNAs and have several members with overlapping functions. Thus, the simultaneous knockdown of all miRNAs in a family can be the best way to confirm the biological function of a specific miRNA family. To fill this lacune, different molecular techniques that knockdown of miRNAs by target decoys/mimics such as target mimics (TMs), short tandem target mimics (STTMs), molecular sponges (SPs) and artificial miRNAs (amiRNAs), have been demonstrated to be useful to underpin the role of these miRNAs. For instance, Franco-Zorrilla et al. (2007) identified an endogenous long non-protein coding gene INDUCED BY PHOSPHATE STARVATION 1 (IPS1) which altered the protein levels of PHOSPHATE2 (PHO2) by modulating the effects of miR399 in Arabidopsis.

It was observed that both IPS1 and PHO2 had highly conserved sequence motifs that contain complementary binding site for the phosphate (Pi) starvation–induced miRNA miR399. However, IPS1 showed three additional nucleotides bases, which provoke central mismatches in the miR399 binding site by forming of central "bulge" opposite the expected miRNA cleavage site, thus avoiding miR399-guided cleavage but instead sequesters of miR399. This endogenous regulatory mechanism of miRNA activity was termed as endogenous "target mimicry" or eTM which is commonly referred to as miRNA decoys/mimics technology. Based on the strategy of mimicking target transcripts, new methods to inhibit miRNAs, particularly Short Tandem Target MIMICs (STTMs)-based knockouts, have been developed in miRNA functional studies (Tang et al., 2012; Teotia et al., 2016). STTMs contain two miRNA-

binding sites, which not only efficiently silence some highly abundant miRNAs but also can be used to study the interactions between two miRNAs by inserting two different miRNAs in the same STTM construct (Zhang et al., 2017; Peng et al., 2019).

In the area of *in silico* miRNA target prediction in plants, numerous bioinformatics tools can be used as web servers and freely accessed including TAPIR (https://ptarpmir.cu-bic.ca/), TarHunter (http://www.biosequencing.cn/TarHunter/), psRNATarget (https://www.zhaolab.org/psRNATarget/), psRobot (http://omicslab.genetics.ac.cn/psRobot/) and WPMIAS (https://cbi.njau.edu.cn/WPMIAS/), these three last lacking miRNA-initiated phasis information (Liu et al., 2021b). Furthermore, PHASIS (https://github.com/atulkakrana/PHASIS) and PhaseTank (https://phasetank.sourceforge.net/) were developed to support phasiRNAs prediction in plants. The database sRNAanno is presented as a comprehensive collection of phasiRNA loci in plants (http://plantsrnas.org/), but still, does not indicate which phasiRNA sites are triggered by miRNAs. More recently, TarDB (http://www.biosequencing.cn/TarDB/) surged as a web resource for exploring relatively high-confidence miRNA targets and miRNA-triggered phasiRNAs in plants RNA information (Liu et al., 2021b). An example of such specific miRNA targeting use can be exemplified by the recent 216 drought-responsive identified miRNAs (DRMs) based on 28 drought stress-specific sRNA datasets using the RiceMetaSys: Drought-miR database (Kumar et al., 2024). At a genome-wide scale, the high throughput degradome/PARE (Parallel analysis of RNA ends) sequencing techniques have enabled the characterization of miRNA cleavage sites. Computational pipelines such as CleaveLand, PARESnip and sPARTA can help further analyse the data sets originated through degradome/PARE-seq (Addo-Quaye et al., 2009; Kakrana et al., 2014; Folkes et al., 2012). Databases for plant miRNA such as miRBase (https://mirbase.org/) and PmiREN (https://www.pmiren.com/) can be utilised not only to archive and annotate miRNAs but also as a miRNA encyclopaedia with function and resource tools.

6.5 Precision gene editing tools for crop improvement

Over the past three decades, the unconditional success of genetically modified (GM) crops allowed the way for harnessing novel plant breeding technologies in crop improvement. Despite the widespread adoption of GM crops, the presence of foreign genes in them since their development has been central concern of public acceptance and potential biosafety issues in many parts of the world (Kumar et al., 2020). These concerns have driven the development of new plant breeding techniques, such as gene editing, which enable precise modification of plant genomes without introducing foreign DNA (Wang and Doudna, 2023).

The foundation of gene editing in plant was established with the discovery of I-SceI (meganuclease) induced double stranded breaks in a site-specific manner that enhanced homologous recombination in plants (Puchta et al., 1993). Meganucleases, also referred to as homing endonucleases, are sequence-specific endonucleases which found in prokaryotes, archaea, and unicellular eukaryotes. These endonucleases are often small proteins that are encoded by mobile genetic elements, either introns or inteins, and cleave relatively long sequences of DNA (ranging from 12-40 bp) resulting in higher specificity and lower off target cutting (Stoddard, 2014). The primary challenge facing meganuclease technology is the significant engineering efforts required to create enzymes with novel DNA specificity, particularly due to the overlapping nature of DNA-cleavage and DNA-binding domains in these homing endonucleases. Despite this hurdle, engineered meganucleases have demonstrated some success in various gene editing functions in plant species (Daboussi et al., 2015); however, the production of modified homing endonucleases is both time-consuming and lacks the necessary flexibility.

To address the problem associated with meganucleases, another class of site-specific nucleases (SSN) such as zinc finger nucleases (ZFNs) were developed by fusing an artificial DNA-binding domain (DBD) of a versatile class of eukaryotic transcription factors – zinc finger proteins (ZFPs) to the non-specific DNA cleavage domain of type II restriction endonuclease *FokI* (Kim et al., 1996). This achievement spawned from the discovery that the DNA-binding domain and the cleavage domain of the *FokI* function independently of each other (Li et al., 1992). The FokI nuclease operates as a dimer, thus necessitating the design of two ZFN molecules to target a single site. Each ZFN molecule binds itself to opposite strands of DNA molecule, causing a double-strand break. One notable development is that they have modified a vast variety of genomes in plants (Urnov et al., 2010). Yet, the only thing standing in the way is its challenges such as off-target cutting, and cytotoxicity which led to the development of a new class of nucleases: transcription activator-like effector nucleases (TALENs). TALENs, akin to ZFNs, arise from the fusion of catalytic domain of the FokI endonuclease with a DNA-binding domain called transcription activator-like effectors (TALEs), derived from the protein produced by plant pathogenic *Xanthomonas* spp. Bacterium (Christian et al., 2010; Bogdavone and Voytas, 2011).

Until a few years ago, both ZFNs and TALEN were the most powerful programmable site-specific technologies for genome engineering. These techniques are modular in form and function, comprised of DNA-binding domain fused to *FokI* cleavage domain that recognize target sequences by protein motifs and which requires the assemblance of specific proteins to each target, therefore having low specificity in recognizing and cleave the DNA targets (Gaj et al., 2016). Recently, clustered regularly interspaced short palindromic repeats/CRISPR-associated nuclease 9 (CRISPR/Cas9) techniques have been proved to be useful to genome editing because CRIPSR/Cas system has the advantage of recognizing targeted genomic sites by base complementarity between the single-stranded guide RNA (ssgRNA) and the target DNA. So far, CRISPR/Cas9 has been applied to introduce specific genetic modifications aimed at improving desirable traits in cereals, vegetables (including root and tuber crops) and fruit trees (Osakabe and Osakabe, 2015; Rodriguez-Leal et al., 2017; Jaganathan et al., 2018; Tripathi et al., 2020; Biswas et al., 2021; Ma et al., 2023).

Nonetheless, it is important to note that CRISPR/Cas is not a universal tool, as it has limitations related to the short sequence

length and the specificity of the microRNA (MIR) genes. These limitations affect the design of guide RNAs (gRNAs) targeting MIR genes and their applicability, respectively. Precise genome editing requires accurate gene sequences and knowledge of their functions, which is facilitated by the whole-genome sequencing data currently being generated. Transgenic plants overexpressing cleavage-resistant targets such as miR164-resistant *SlNAM2* (Hendelman et al., 2013) in tomato, and miR139-resistant *OsTCP21* (Zhang et al., 2016a) and miR166-resistant *OsHB4* (Zhang et al., 2018a) in rice, have been created and used to investigate miRNAs functions. Among the various loss-of-function systems available, miRNA decoy techniques [Target Mimic (TM) and Short Tandem Target Mimic (STTM)] and the CRISPR/Cas9 genome-editing system are the most utilized. As an example, the resistance to powdery mildew in barley accessions and mutant types was shown to be controlled by the *mlo* gene (the wild allele is *Mlo*) (Jørgensen, 1992). The loss of function of *mlo* mutants confers a durable and broad-spectrum resistance to powdery mildew. Yet, the *mlo*-associated resistance brings along growth and yield penalties, where the expression of *Mlo* in *mlo* genotypes is sufficient to confer single-cell susceptibility. Pleiotropic effects may explain this phenomenon, and many breeding programs may need to address it. In polyploid crops, this technique may require additional backcrossing cycles to eliminate any unwanted mutations (Figure 6).

In hexaploid wheat, the TALEN-generated Tamlo-R32 mutant targets the three wheat *MLO1* genes (*TaMLO-A1*, *TaMLO-B1* and *TaMLO-D1*) and is characterised by a 300kb pair targeted deletion in the *MLO-B1* locus that retains crop growth and yields while conferring resistance to powdery mildew. Through epigenetic changes, this mutant exhibits a significant upregulation of *TaTMT3B* (Tonoplast monosaccharide transporter 3) locus transcript expression. The use of the CRISPR/Cas9 tool demonstrated that this genetic arrangement counteracts the negative effects associated with *mlo* mutations while maintaining strong disease resistance in wheat (Li et al., 2022).

In strawberry, *fvesep3* mutated using CRISPR/Cas9 has produced the desirable trait of parthenocarpic fruits, which is highly sought in strawberry breeding programs (Liu et al., 2020b; Pi et al., 2021). Similarly, in polyploid and parthenocarpic bananas, where asexual breeding methods are less efficient, the utilization of CRISPR/Cas9 systems can enhance mutation efficiency and introduce traits of economic importance such as reduced plant height (Shao et al., 2020), improved quality (Kaur et al., 2020; Awasthi et al., 2022), and host plant resistance to pathogens (Tripathi et al., 2021).

Recently, and in response to some constraints presented by Cas9, Zheng et al. (2024) demonstrated in rice that CRISPR Cas12a is a more efficient tool compared to its Cas9 counterpart for generating knockout mutants of a miRNA gene. With this improvement, it seems possible to achieve editing efficiencies of up to 100%. Due to its ability to create larger deletions, which facilitates the generation of loss-of-function mutants of targeted genes, this new system appears better suited for developing genome-wide miRNA mutant libraries, at least in rice.

Improving crops by pyramiding of a few desired genes in a single elite line have been facilitated by marker-assisted selection (MAS). As shown for diverse crop species in the following sections, CRISPR take this a step further by the so-called multiplexing of a number of genes that have been mutated by CRISPR/cas in a single event. In fact, a mutant collection can be generated by use of CRISPR as demonstrated for maize in the BREEDIT pipeline by Lorenzo et al. (2023) where a first attempt of multiplexing 48 growth-related genes resulted in more than thousand mutant lines which can be screened for improvement in complex traits such as drought and heat tolerance.

6.6 Case studies

6.6.1 Rice

In rice, noteworthy examples of precision editing include the knockout of three seed weight-related genes, namely GW2, GW5, and TGW6, which lead to an increase in grain weight (Xu et al., 2016). In this crop, the targeted disruption of the promoter region of the SWEET gene has been shown to enhance rice's resistance to bacterial blight (Oliva et al., 2019). In quest to boost African rice's agronomic potential, a method known as multiplex CRISPR/Cas9 was used to target *HTD1*, *GS3*, *GW2*, and *GN1A* loci, which bear domestication genes for the cultivation of rice. It has been established that mutations in these genes can lead to significant increase in seed output when tested in landraces of rice (Lacchini et al., 2020). In a different experiment, the *GS3* locus was targeted alongside with the *qSH1*, *An-1*, and *SD1* loci, that all together led to rapid domestication of wild allotetraploid rice (Yu et al., 2021). These examples highlight the effectiveness of multiplex CRISPR/Cas9 technique in rice that can lead to improve and advance crop domestication efforts.

6.6.2 Wheat

In wheat, CRISPR/Cas-mediated genome editing had been targeted traits such as host plant resistance to powdery mildew (*TaMLO* & *TaEDR1*) (Shan et al., 2013; Wang et al., 2014; Zhang et al., 2017), seed yield (*TaGW2* & *TaGASR7*) (Zhang et al., 2016, 2018b; Wang et al., 2018a), fusarium head blight resistance (*TaHRC*) (Su et al., 2019), herbicide resistance (*TaALS* & *TaACC*) (Zhang et al., 2019a, 2019), quality traits such as low gliadin (*VIT2*), starch and amylose (*TaSBElla*) contents (Jouanin et al., 2019; Liu et al., 2020; Li et al., 2020b), high haploid induction rate (*TaCENH3α*) (Lv et al., 2020), male sterility (*TaMs1 and TaNP1*) (Okada et al., 2019; Li et al., 2020c), increase of P uptake (*TaPHO2-A1)* (Ouyang et al., 2016) and storability (*TaLOX2*) (Shan et al., 2014; Zhang et al., 2016). These modifications were performed by using edits such as knockout, multiplexing, base editing, primer editing and HDR replicon (Li et al., 2021). In addition, the knockout of numerous conserved domains in approximately 100 α-gliadin family members facilitated the development of a low-gluten cultivar suitable for individuals with celiac disease (Sánchez-León et al., 2018). The existence of three sub-genomes (A, B, and D) in wheat

makes the application of multiplex genome editing beneficial, although achieving precise editing remains challenging as current efforts predominantly focus on random mutations and knockouts through non-homologous end joining (NHEJ) repairing. In crops like wheat NHEJ is commonly exploited to induce these random changes in the DNA sequence, allowing researchers to disrupt or deactivate targeted genes for various purposes.

6.6.3 Solanum genus

In tomato (*Solanum lycopersicum* L.), the multiplexing of the coding regions of *SELF-PRUNING* and *SELF-PRUNING 5G*, together with cis-regulatory regions of *CLV3* and *WUS* or open reading frames (ORFs) of GGP1, allowed the generation of tomato fruits with compact plant architecture, synchronized fruit ripening, day length insensitivity, enlarged fruit size and increased vitamin C levels (Rodriguez-Leal et al., 2017; Li et al., 2018). Additionally, leveraging the multiplex capacity of CRISPR/Cas9, researchers successfully targeted five different genes associated with lycopene content accumulation in tomatoes (Li et al., 2018). More recently, in potatoes (*Solanum tuberosum* L.), the functional knockouts of several *S*-genes, namely *StDND1, StCHL1*, and *DMG400000582* (*StDDMR6-1*), lead to the generation of tetraploid potatoes with increased resistance against late blight (Kieu et al., 2021). The mutation of *DMR6* homologues indicates that *StDMR6-1* and *StDMR6-2* have two different biological functions, with the first involved in pathogen resistance whereas the second involved in plant growth. It was observed that *Stdnd1* and *Stchl1* mutants reduced infection lesion sites, whereas StDMR6-1 not only reduced the infection lesion sites but also the percentage of infected leaves. These results are very promising for the potato breeding since pathogen races in potatoes rapidly overcome the existing plant resistance.

6.6.4 Seed production

CRISPR/Cas9 offers a potential solution to avoid the need for farmers to yearly seed purchase due to the phenomenon of heterosis. Heterosis, also known as hybrid vigour, is only maintained within the F_1 generation, which means that the production of these superior offspring is laborious, costly, and often unaffordable to small scale farmers. Nevertheless, if CRISPR/Cas9 is used then it can help to overcome such limitations and allowing for the development of stable and improved seed cultivars that retain desirable traits across generations. This would mean that farmers would have a reducing dependence on purchasing new seeds every year, and, therefore, providing a more sustainable and cost-effective solution for farmers.

Through simultaneous editing of *REC8*, *PAIR1*, *OSD1*, and *MTL* genes, researchers have been able to fix the favourable F_1 traits (Wang et al., 2019). When *REC8*, *PAIR1*, and *OSD1* genes were knocked down simultaneously, and meiosis was replaced by mitosis, it was enabled the production of asexual hybrid rice seeds and the preservation of the hybrid vigour (Khanday et al., 2019). In strawberries, and other fruits crops, the knockdown of all the six anthocyanin transport genes *RAP* (*REDUCED ANTHOCYANINS IN PETIOLES*) using the CRISPR/CAS9 system has produced white octoploid fruits instead of the red-coloured typical berries (Gao et al., 2020). This demonstrates the simultaneous knockout of homoeologous alleles as a tool to breed polyploid plants.

Similarly, the generation of DH (double-haploid) homozygous lines is a labour-intensive and expensive endeavour that needs laboratory infrastructure and has variable efficiency. The timeline for the whole DH line production process takes from several months to a few years depending on factors such as the species involved, the haploid induction success rate and the chromosome doubling technique efficiency. However, targeted genome editing-mediated haploid induction has provided a more rapid option with lines being generated within one year (Hooghvorst and Nogués, 2020). Several key genes have been targeted for this purpose including *MATL* (*MATRILINEAL*), *CENH3* (*CENTROMERE-SPECIFIC HISTONE 3*) and *DMP* (*Domain Membrane Proteins*) genes (Zhu et al., 2020; Kuppu et al., 2020; Zhong et al., 2019a). In wheat CRISPR/Cas9-mediated targeting of the *MATL* gene led to an inheritance rate of 18.9% haploid progeny (Liu et al., 2019a, b) while targeting maize *DMP* gene resulted in maternal haploids with an efficacy range of about 0.1–0.3% (Zhong et al., 2019a, b). An improved version of this system is the HI-editing technology (HI-Edit) which makes available transgene-free edited inbred lines lacking inducer parental DNA. By eliminating the necessity for introgression, this approach minimizes the time and expense involved (Kelliher et al., 2019). As technological advancement carries on, techniques like the CRISPR-Combo system which performs simultaneous gene activation and editing using CRISPR hold a lot of promise in terms of speeding up the identification process of markers-free genome-edited lines that are highly efficient. These techniques also offer an opportunity to screen mutants at both genome and transcriptional levels (Pan et al., 2022a). As these technologies continue to evolve, it's expected that the process of obtaining and studying gene-edited lines will become more efficient and streamlined. Hence, we can look forward to easier and faster ways of getting and understanding these edited genetic lines in the future.

7 CRISPR/Cas9 without DNA donor involvement

Ever since the first mention of CRISPR/Cas9 technology, back in 2013 (Li et al., 2013; Nekrasov et al., 2013; Shan et al., 2013) there have been some major improvements. In the past, this technology relied on using a DNA donor template to introduce the desired genetic changes. However, the most recent developments have allowed for the utilisation of CRISPR/Cas9 without the need for a DNA donor template, opening new possibilities for gene editing. This modified version of CRISPR/Cas9, instead of using a DNA donor, uses the Cas9 enzyme to produce targeted breaks in specific parts of the genome. These breaks then activate the cell's natural DNA repair mechanisms, which can result in gene disruptions,

insertions, or deletions. The CRISPR/Cas9 technology involves the cut of a specific sequence in the organism's genome. This cutting action is triggered by an RNA called single-guide RNA (sgRNA), which is designed to match the target DNA sequence. Double-stranded breaks are generated when the homologous sequences of the guide RNA are in proximity to the specific protospacer adjacent motif (PAM) sequence. This allows for genetic modifications to occur during the subsequent repair mechanisms (Asmamaw and Zawdie, 2021; Singh et al., 2022). More recent advancements in CRISPR technology made it possible to combine an engineered reverse transcriptase and prime editing (PE) RNA with CRISPR/Cas9. This new approach eliminates the need for double-stranded breaks or donor templates and enable changing specific bases, deleting targeted sections, inserting, or even a combination of those. This methodology has been demonstrated on crops such as maize, lettuce, tobacco, rice, and bread wheat, and is a potential solution to tackle concerns raised by the public and regulatory bodies about CRISPR/Cas9-derived plants (Svitashev et al., 2015; Woo et al., 2015; Liang et al., 2017).

8 CRISPR/Cas9 prime-editors in plants

CRISPR/Cas9 prime-editors have become a powerful tool for precise editing of plant genomes. Prime-editing combines the Cas9 nuclease with an engineered reverse transcriptase and a prime editing RNA (pegRNA), which enables targeted base editing, insertions, deletions, and combinations without the need for double-stranded breaks or donor templates. Currently, the most usual editors used in the CRISPR system are the base editors, and characterised for making specific base changes, limited to A, C, G, and T. Instead, prime editors (PEs) offer a broader range of mutations, including all 12 possible transition and transversion mutations, as well as small insertions and deletions (INDELs), without the need for DNA double-strand breaks (Anzalone et al., 2019). One example of prime editing in plants is the modification of the acetyl-CoA carboxylase (*ACCase*) gene. This gene is crucial for fatty acid biosynthesis, and the loss-of-function mutations on it can lead to severe developmental abnormalities or even be lethal for plants (Baud et al., 2004). *ACCase* is also the target of many commercial herbicides. In rice, the *OsACC* gene confers herbicide resistance. Through base editing libraries, targeted mutations were introduced in both *OsACC* and acetolactate synthase (*ALS*) genes, which are targets of herbicides used for plant weed control in field crops (Garcia et al., 2017). Using this approach key amino acid residues that are related to resistance to herbicides were revealed (Liu et al., 2020; Kuang et al., 2020). In addition, various mutations in the acetyl-coenzyme A carboxylase 1(*OsACC1*) gene, have been discovered, resulting in herbicide resistance (Xu et al., 2021). These examples illustrate the significance of prime editing technology for crop improvement purposes specifically when it comes to development of herbicide-resistant cultivars and understanding the functional importance of specific substitutions of amino acids. Although its optimization is still underway, prime editing in plants promises precise genetic manipulation and therefore serves as a promising tool towards better agricultural achievements.

9 Opportunities and obstacles in applying CRISPR/Cas9-derived methodologies

The access to complete DNA sequencing information and a thorough understanding of gene functions has facilitated the use of CRISPR-Cas9 technology for precise modulation of key genes. This also allows for the quick creation of new genetic resources to improve specific traits. Nevertheless, it's crucial to note that direct sequencing methods might miss out on certain heterozygous alleles. To overcome this limitation, methods like TILLING (Targeting Induced Local Lesions IN Genomes) can be used to identify mutations at specific loci in the modified genetic material. Furthermore, DNA markers can be used to detect mutations at the early stages of plant growth. TILLING brings several benefits in this scenario. Libraries developed through TILLING contain preserved genetic resources, usually in the form of seeds, and the genomic DNA extracted from mutated materials can remain stable for many years. This allows for the identification of specific allele groups linked to different traits, which can be safeguarded and shared with the public. However, caution is advised when targeting traits controlled by multiple genes. The current landscape provides numerous prospects for breeders and scientists, depending on their gene preferences. When dealing with stable genes, hybridization is often the preferred strategy. Conversely, if the genes of interest are dominant, transformation approaches may be the most effective option. For recessive genes, mutation methodologies are frequently utilized. These strategies, based on our existing knowledge, are implemented in plant research to tackle urgent global challenges (Figure 7).

By employing refined NGS pipelines (Sahu et al., 2020) to characterize mutants generated through diverse chemical and irradiation techniques, new allelic variants in the genomes of various crops can be generated. This approach potentially may establish mutation breeding as a fundamental pillar in the efforts towards nutritional and food security. It is also an opportunity for the generation, validation, and removal of deleterious alleles, since these are mainly accumulated in the pericentromeric regions of chromosomes due to less selection efficiency (Dwivedi et al., 2023). Yet, one must keep in mind that these regions house a great amount of "housekeeping genes" that could produce deleterious effects if disturbed. Additionally, these regions might be protective of these types of genes. Whereas the CRISPR process transforms and introduces non-native genes, it can also generate transgene-free derivatives through genetic segregation via self-pollination. This is achieved by providing a fluorescent cassette marker genetic segregation, which signals the presence of the CRISPR/Cas9 construct. Another approach is to introduce suicidal genes (CMS2 and BARNASE genes) into the T_0 plants, which selectively eliminate transgene-containing pollen and embryos through the crossing. These methods are, however, still not applicable for asexually propagated crops like potato, cassava, and banana. It's important to note that while CRISPR is often associated with simple mutations, it can also facilitate the introduction of larger foreign DNA fragments, which is categorized as NGT 2 in the EU's soft regulation framework (EC, 2023). This distinction is important, as it

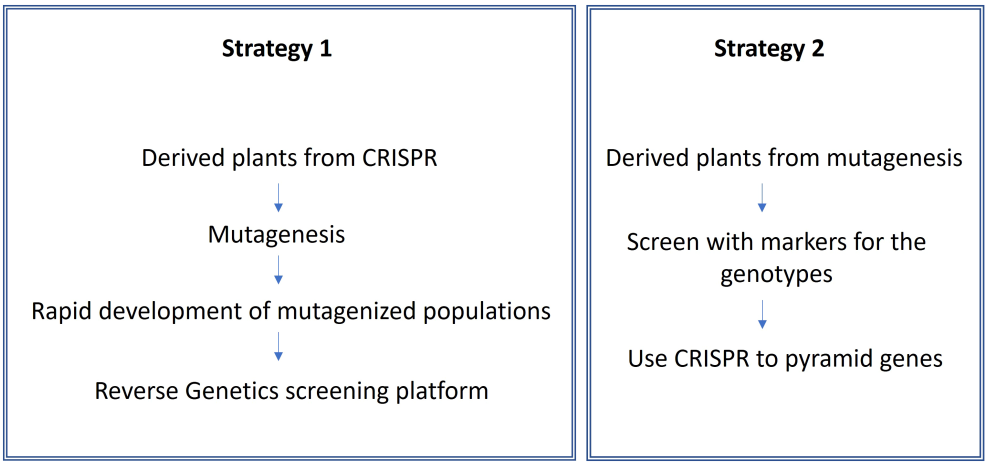


FIGURE 7
Strategies on how CRISPR and mutagenesis can be applied based on current knowledge.

highlights that CRISPR's capabilities extend beyond short, simple mutations typically seen with chemical-induced mutations, and therefore, enabling more complex genetic modifications.

Among the known challenges, it is worth mentioning the difficulty in predicting sgRNAs efficiencies, requiring the optimization of Cas protein derivates. The current delivery systems for Cas and sgRNAs have limitations, highlighting the need for improvement in the CRISPR specificity system, and the urgent need to expedite improvements in these delivery mechanisms (Table 3).

The abundance of genomic sequence data and the need to uncover target genes associated with essential agronomic traits highlights the importance of establishing precise links between genotypes and phenotypes. Therefore, better delivery systems in CRISPR technology are needed, including reducing the dependency on tissue culture methods and to facilitate the widespread adoption of this technology in agriculture.

While the technology seems useful for specific traits such as plant diseases and pests, or quality traits; when it comes to more complex traits, such as yield, the potential of CRISPR technology may be seen with some reservations. CRISPR technology seems to be a valuable tool to facilitate trait introgression and support the mitigation of linkage drag in the case of simply inherited traits. However, in both CRISPR and induced mutation, there is the need to address challenges of specific genetic variance generation for highly quantitative traits, which is crucial to achieve the necessary increases in production. We are likely to experience great innovations in crop genetics in a multitude of ways, yet, if accurate phenotyping will not accompany this evolution, it will be challenging to observe a rapid evolution in crop improvement. Integrating precise genotyping with accurate phenotyping will significantly contribute to the effective implementation of CRISPR technology and accelerate the progress in crop improvement efforts.

Regarding regulatory frameworks, currently, there are two regulatory frameworks that several countries have adopted to classify crops developed using CRISPR and other genome editing technologies: product-based and process-based regulations (Ishii and Araki, 2017). In the product-based framework, only the final product is controlled in terms of regulation. Consequently, if there are no exogenous transgenes in the commodity, then it is considered as a non-GMO. Conversely, under a process-based approach, any crop produced through CRISPR/Cas9 technology shall be considered a genetically modified organism (GMO). The choice between these two frameworks will determine how genome-edited crops should be managed and labelled. Different countries have adopted different approaches, and specific examples of countries following each framework can be found in related literature (Chen and Gao, 2020; Singh et al., 2022). Regulatory setups significantly influence commercialization as well as public reception of CRISPR/Cas9-derived crops across various regions globally. Despite technologies such as ZFN, TALEN, Meganucleases (MN), and CRISPR can be used for specific genome editing without the incorporation of foreign DNA into the target site, it is crucial to recognize that these tools are not perfect and sometimes irregularities occur during the process. The off-target effects are one major worry because they may result in cell death or cause genomic instabilities (Lazzarotto et al., 2020) while sequencing cannot detect or differentiate between natural and genetic engineered variability. Research demonstrates that prime editors (PEs) do not induce detectable pegRNA-independent off-target edits in plants (Jin et al., 2021). Using CRISPR to treat genetic diseases in humans, side effects such as chromosomal rearrangements can occur after genome editing (Raimondi et al., 2024), besides editing a gene may turn on or off other genes associated with that gene. In plants, off-targets and side effects may be segregated out during meiosis or discarded during the multigenerational field evaluation selection because of performance and therefor side effects may be of lesser practical problem in seed propagated plants. For asexually propagated plants, field assessments may be the only way of recognizing such effects. Nevertheless, continued research in the field of genome editing is needed to address these concerns and ensure the safe and precise

TABLE 3 Challenges and options to overcome in the CRISPR/Cas9 systems delivery.

System	Challenge	Possible solutions	Reference
CRISPR/Cas9	Editing efficiency	Increase CG contents in the sgRNA by selecting sgRNA with high CG content	Ren et al. (2019)
		Use of native U6 promoters	Long et al., 2018; Zhou et al., 2018; Ren et al., 2021; Zhang et al., 2022
		Use of Cas12a/b proteins to extend CRISPR usage scope	Ming et al., 2020
		Use of fluorescent labelling	Ali et al., 2023
	Multiplex Editing	Usage of PTG/Cas9 system	Wang et al., 2018b
		Use of multiplexed tRNA-gRNA2.0 system	Pan et al., 2021
	Deliver efficiency into plant cells	Use of vector-mediated and nanoparticle delivery systems; use of cut-dip-budding delivery system in herbaceous and woody plants; use of merismatic and plant germlines	Laforest and Nadakuduti, 2022; Wang et al., 2022b; Cao et al., 2023; Ali et al., 2023
	Off-target effects	Use of engineered precision variants of Cas9, Cas12a and deaminases; use high-fidelity Cas9	Zhang et al., 2019; Jin et al., 2020a; Chen et al., 2017
	Side effects	Segregation, field evaluation	Discussed below (This work)
	Low regeneration rates lead to high chimerism	Use of adventitious regeneration protocols; use of chimeric genes	Malabarba et al., 2020; Pan et al., 2022b
	Generation of transgene-free plants	Use of lipid transfection, viral vectors, delivery of components directly as functional sgRNA and Cas9 protein	Mahmoud et al., 2022; Ma et al., 2020; Pompili et al., 2020
	Activation of gene expression	Stabilization of donor DNA and CRISPR system through the introduction of 5'-phosphorylated double-stranded oligodeoxynucleotide (dsODN)	Lu et al., 2020
	Precision	Use of base editor tool	Gao, 2021
		Improve the efficiency of the primer editor for usage	Chen et al., 2021; Nelson et al., 2022
	Epigenetic modifications	Use of more efficient epigenome editors	Wilson et al., 2020
	Use of double-haploid technology	Higher production of haploids	Zhong et al., 2019a; Kuppu et al., 2020; Gao, 2021
	Regulations	Non-integration of foreign DNA/RNA into the target site	Wang et al. (2022c)

application of these technologies in economically important crops. For small breeders that obtain royalties for their varieties, or payment for certified seeds and farmer saved seeds, outsourcing CRISPR technology is a good option for them to get access to specific traits without investing in costly laboratory facilities and staff with specialised skills.

10 Connecting the dots – Fostering the integration of inclusive genomic innovations in the smallholder farmers' agritech investments

The successful adoption and implementation of inclusive genomic innovations (IGIs) in the agriculture sector of the Global South, especially in low- and mid-income countries, continue to pose challenges. The primary obstacle stems from a lack of understanding and awareness regarding the specific needs and requirements of the target users (UNDP, 2022).

For instance, during the late 1990s, cotton crop was significantly affected by insecticide-resistant Lepidopteran pests in India resulting in excessive use of insecticides that caused not only a serious environmental and human health problem but also led to economic problems for cotton growers that contributed to farmer indebtedness. The introduction of genetically modified cotton [*Bacillus thuringiensis* (Bt) cotton] showed not only a direct positive impact on yield gains through reduced crop losses by good control of native bollworm *Helicoverpa armigera*, but indirectly also led to reductions in insecticide use (Gutierrez et al., 2020) resulting in 134% increase the income of smallholder farmers (Subramanian and Qaim, 2010). In China, the widespread planting of Bt cotton cultivars not only successfully controlled the effects of the polyphagous pest *Helicoverpa armigera* among multiple crops but also provided great advantages by providing a 10-fold increase in the products of Chinese farmers between 1996–2018 (Pray et al., 2011; Zhaozhi et al., 2022; ISAAA,

2020). Similarly, the adoption of Bt technology in other crops such as Bt brinjal contributed 22% higher revenues compared to non-Bt cultivars in Bangladesh (Shelton et al., 2020).

Despite these encouraging outcomes, public perception, concerns about potential risks and unintended consequences, and ethical considerations contribute to the ongoing debate. An alternative approach includes the use of 'Fast Identification of Nucleotides variants by DigITal PCR' known as FIND-IT, which offers a promising solution for rapid identification of pre-targeted genetic variants or rare alleles in large and very large populations. With this method libraries of 500,000 knockout barley mutant individuals, can be screened within only two weeks (Knudsen et al., 2021; Madsen et al., 2024). Contrary to CRISPR methodology, FIND-IT is not subjected to governmental regulations as a non-GM technique. Additionally, compared to TILLING methods, FINF-IT offers simpler technological requirements and greater sensitivity in detection. This high-throughput ddPCR method does not require transformation or tissue culture protocols making it a scalable and efficient approach for screening and targeting desired traits in crops with low mutation-density variant populations.

The innovative upfront genomic innovation (GI) technologies may not directly serve as tools for smallholder farmers in many countries but rather offer products that can benefit them. It is undeniable that smallholder farmers are a significant portion of the global agricultural workforce, operating, majority of the times, under resource constrains. Therefore, investments in high technologies such as genotyping are opportunities to improve yields, reduce production costs and even mitigate risks associated with the increased climate variability. The solution is to bridge the gap that exists between traditional farming methods and modern agricultural practices. By offering them tailored solutions to address their unique challenges, genomic innovations may make a difference. The information got from the genetic codes of crops, not only offers better cultivars that are better locally adapted but also facilitates the selection of the most desired traits and allows the farmers to produce higher-quality crops with fewer resources. Yet, to fully understand the potential of genomic innovations requires addressing issues such as access, affordability and capacity building to ensure equitable distribution of benefits across the farming communities. Besides the previous examples, in Sub-Saharan Africa, the development of drought-tolerant maize cultivars using MAS breeding techniques using genomic data (molecular markers) has helped farmers mitigate the adverse effects of climate change and improve food security (https://dtma.cimmyt.org/). The same was shown by the introduction of disease-resistant cassava cultivars in Nigeria and Uganda (https://www.nextgencassava.org/). With an exponential amount of information available, the future holds the opportunity to achieve tailored genomic technologies for smallholder farming. Such examples of this are MAS, GS and genome editing tools. One way to foster the integration of inclusive genomic innovations into agritech investments for smallholder farmers is to prioritize capacity building and knowledge transfer, including through training programs, extension services and farmer field schools.

Innovative financing mechanisms, including impact investing, venture capital and blended finance models may support resource mobilization and scale up the adoption of inclusive genomic innovations and therefore, accelerate their adoption by the farmer community. An example of this is the effort on biofortification through breeding by the HarvestPlus program (www.harcestplus.org) that over 20 years facilitated the release of more than 420 biofortified cultivars in different staple crops across Asia, Africa and Latin America (Dwivedi et al., 2023). So far, most of the biofortified cultivars developed through crossbreeding approaches, however, molecular breeding efforts using IGIs for the same purpose are gaining momentum, thus resulting in speed up the development of biofortified cultivars (Badu-Apraku et al., 2018; Sheoran et al., 2022). The molecular markers developed for biofortified traits such as pro-vitamin A, Fe and Zn content in edible parts of different staple crops could be utilized to improve locally adapted farmer-preferred cultivars in sub-Saharan Africa and Latin America.

Despite all the above-mentioned strengths in empowering smallholder farmers through genomic innovations, careful consideration must be given to potential risks including genetic erosion, biodiversity loss, and unintended environmental consequences. One must be aware that, beyond technology adoption, it is important to take into consideration that the use of CRISPR ultimately raises important concerns regarding Intellectual Property Rights (IPRs), particularly as these technologies can impact breeders' rights, the ownership of genetically edited crops but also issues on the accessibility of the developed innovations in agricultural biotechnology. Potential solutions could involve not only the development of clear legal frameworks, encourage the open access to CRISPR technology findings to the breeders, develop fair and flexible licensing agreements so breeders can use CRISPR technology while providing compensations to innovators, improve communication and collaborations between the different stakeholders, as well promotion for fairer regulations that protect both innovators and breeders.

The future holds space not only to explore genome editing tools for unprecedented precision and efficiency but also the space to integrate the omics data technologies that promise deeper insights into gathering real-time data for real-time decision-making. A major contribution of these advanced technologies, from the breeder perspective, includes the implementation of multiple field trials, which should be emphasised for mutational and CRISPR-Cas generated parental lines after they have demonstrated success in greenhouse conditions.

To conclude, the integration of various approaches, whether it be through mutation breeding, gene editing, crossbreeding or other techniques, holds the potential for a more comprehensive and effective strategy in enhancing plant traits for increased production.

Author contributions

AG: Writing – original draft, Writing – review & editing. RO: Writing – review & editing. FS: Writing – review & editing. SR: Writing – review & editing. PA: Writing – review &

editing. AA: Writing – review & editing. MK: Writing – review & editing. SC: Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. Bill and Melinda Gates Foundation for their support through the RTB Breeding Investment Project (INV-041105).

Acknowledgments

A.L.G.-O. would like to express gratitude to Dr. Emily Conley and Professor Jim Anderson from the Department of Agronomy and Plant Genetics at the University of Minnesota (UoM), Minnesota, U.S.A. for their valuable insights and support in this review. We also extend our appreciation to the *Bill and Melinda Gates Foundation* for their support through the RTB Breeding Investment Project (INV-041105).

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

References

Abe, T., Kazama, Y., and Hirano, T. (2015). Ion beam breeding and gene discovery for function analysis using mutants. *Nucl. Phys. News* 25, 30–34. doi: 10.1080/10619127.2015.1104130

Abe, T., Ryuto, H., and Fukunishi, N. (2012). "Ion beam radiation mutagenesis," in *Plant mutation breeding and biotechnology*, eds. Shu, Q. Y., Forster, B. P., and Nakagawa, H. (Vienna: Joint FAO/IAEA), 99–106. doi: 10.1079/9781780640853.0099

Addo-Quaye, C., Miller, W., and Axtell, M. J. (2009). CleaveLand: a pipeline for using degradome data to find cleaved small RNA targets. *Bioinformatics* 25, 130–131. doi: 10.1093/bioinformatics/btn604

Agre, P. A., Clark, L. V., Garcia-Oliveira, A. L., Bohar, R., Adebola, P., Asiedu, R., et al. (2023). Identification of diagnostic KASP-SNP markers for routine breeding activities in yam (*Dioscorea* spp.). *Plant Gen.* 17, e20419. doi: 10.1002/tpg2.20419

Ahmed, F., Senthil-Kumar, M., Lee, S., Dai, X., Mysore, K. S., and Zhao, P. X. (2014). Comprehensive analysis of small RNA-seq data reveals that combination of miRNA with its isomiRs increase the accuracy of target prediction in Arabidopsis thaliana, RNA. *Biology* 11, 1414–1429. doi: 10.1080/15476286.2014.996474

Ainley, W. M., Dent, S. L., Welter, M. E., Murray, M. G., Zeitler, B., Amora, R., et al. (2013). Trait stacking via targeted genome editing. *Plant Biotechnol. J.* 11, 1126–1134. doi: 10.1111/pbi.12107

Ali, A., Zafan, M. M., Farooq, Z., Ahmed, S. R., Ijaz, A., Anwar, Z., et al. (2023). Breakthrough in CRISPR/Cas system: Current and future directions and challenges. *Biotechnol. J.* 18, 2200642. doi: 10.1002/biot.202200642

Anderson, W., Taylor, C., McDermid, S., Ilboudo-Nebie, E., Seager, R., Schlenker, W., et al. (2021). Violent conflict exacerbated drought-related food insecurity between 2009 and 2019 in sub-Saharan Africa. *Nat. Food* 2, 603–615. doi: 10.1038/s43016-021-00327-4

Anzalone, A. V., Randolph, P. B., Davis, J. R., Sousa, A. A., Koblan, L. W., Levy, J. M., et al. (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* 576, 149–157. doi: 10.1038/s41586-019-1711-4

Arshad, M., Feyissa, B. A., Amyot, L., Aung, B., and Hannoufa, A. (2017a). MicroRNA156 improves drought stress tolerance in alfalfa (Medicago sativa) by silencing SPL13. *Plant Sci.* 258, 122–136. doi: 10.1016/j.plantsci.2017.01.018

Arshad, M., Gruber, M. Y., Wall, K., and Hannoufa, A. (2017b). An insight into microRNA156 role in salinity stress responses of alfalfa. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.00356

Arshad, M., and Hannoufa, A. (2022). Alfalfa transcriptome profiling provides insight into miR156-mediated molecular mechanisms of heat stress tolerance. *Genome* 65, 315–330. doi: 10.1139/gen-2021-0099

Asmamaw, M., and Zawdie, B. (2021). Mechanism and applications of CRISPR/cas-9-mediated genome editing. *Biologics* 15, 353–361. doi: 10.2147/BTT.S326422

Awasthi, P., Khan, S., Lakhani, H., Chaturvedi, S., Shivani, N., Kaur, N., et al. (2022). Transgene-free genome editing supports CCD4 role as a negative regulator of beta-carotene in banana. *J. Exp. Bot.* 73, 3401–3416. doi: 10.1093/jxb/erac042

Badu-Apraku, B., Talabi, O. A., Garcia-Oliveira, A. L., and Gedil, M. (2018). IITA scientists develop multiple stress tolerant maize hybrids with high levels of Pro-Vitamin A. IITA News 2463. 7th january 2019. Available online at: https://bulletin.iita.org (Accessed October 19, 2024).

Baud, S., Bellec, Y., Miquel, M., Bellini, C., Caboche, M., Lepiniec, L., et al. (2004). gurke and pasticcino3 mutants affected in embryo development are impaired in acetyl-CoA carboxylase. *EMBO Rep.* 5, 515–520. doi: 10.1038/sj.embor.7400124

Bento, V. A., Ribeiro, A. F. S., Russo, A., Gouveia, C. M., Carodo, R. M., and Soares, P. M. M. (2021). The impact of climate change in wheat and barley yields in the Iberian Peninsula. *Sci. Rep.* 11, 15484. doi: 10.1038/s41598-021-95014-6

Biswas, S., Zhang, D., and Shi, J. (2021). CRISPR/Cas systems: opportunities and challenges for crop breeding. *Plant Cell Rep.* 40, 979–998. doi: 10.1007/s00299-021-02708-2

Bogdavone, A. J., and Voytas, D. F. (2011). TAL effectors: customizable protein for DNA targeting. *Science* 333, 1843–1846. doi: 10.1126/science.120409

Browning, B. L., and Browning, S. R. (2009). A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *Am. J. Hum. Genet.* 84, 210–223. doi: 10.1016/j.ajhg.2009.01.005

Cao, X., Xie, H., Song, M., Lu, J., Ma, P., Huang, B., et al. (2023). Cut-dip-budding delivery system enables genetic modifications in plants without tissue culture. *Innovation (Camb)* 4, 100345. doi: 10.1016/j.xinn.2022.100345

Chander, S., Garcia-Oliveira, A. L., Gedil, M., Shah, T., Otusanya, G. O., Asiedu, R., et al. (2021). Genetic diversity and population structure of soybean lines adapted to sub-saharan africa using single nucleotide polymorphism (SNP) markers. *Agronomy* 11, 604. doi: 10.3390/agronomy11030604

Chen, J. S., Dagdas, Y. S., Kleinstiver, B. P., Welch, M. M., Sousa, A. A., Harrington, L. B., et al. (2017). Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. *Nature* 550, 407–410. doi: 10.1038/nature24268

Chen, K., and Gao, C. (2020). Genome-edited crops: how to move them from laboratory to Market. *Front. Agr. Sci. Eng.* 7, 181–187. doi: 10.15302/J-FASE-2020332

Chen, P. J., Hussmann, J. A., Yan, J., Knipping, F., Ravisankar, P., Chen, P. F., et al. (2021). Enhanced prime editing systems by manipulating cellular determinants of editing outcomes. *Cell* 184, 5635–5652. doi: 10.1016/j.cell.2021.09.018

Christian, M., Cermak, T., Doyle, E. L., Schmidt, C., Zhang, F., Hummel, A., et al. (2010). Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics* 186, 757–761. doi: 10.1534/genetics.110.120717

CIMMYT (2021). Drought-tolerant maize project pioneers a winning strategy for a world facing climate change. Available online at: https://www.cimmyt.org/news/drought-tolerant-maize-project-pioneers-a-winning-strategy-for-a-world-facing-climate-change/ (Accessed 19 October, 2024).

Curtin, S. J., Zhang, F., Sander, J. D., Haun, W. J., Starker, C., Baltes, N. J., et al. (2011). Targeted mutagenesis of duplicated genes in soybean with zinc-finger nucleases. *Plant Physiol.* 156, 466–473. doi: 10.1104/pp.111.172981

Daboussi, F., Stoddard, T. J., and Zhang, F. (2015). "Engineering meganuclease for precise plant genome modification," in *Advances in new technology for targeted modification of plant genomes*. Eds. F. Zhang, H. Puchta and J. Thomson (Springer, New York, NY). doi: 10.1007/978-1-4939-2556-8_2

Djami-Tchatchou, A. T., Sanan-Mishra, N., Ntushelo, K., and Dubery, I. A. (2017). Functional roles of microRNAs in agronomically important plants - potential as targets for crop improvement and protection. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.00378

Donald, C. M. (1968). The breeding of crop ideotypes. *Euphytica* 17, 385–403. doi: 10.1007/BF00056241

Duhan, N., and Kaundal, R. (2021). LegumeSSRdb: A comprehensive microsatellite marker database of legumes for germplasm characterization and crop improvement. *Int. J. Mol. Sci.* 22, 11350. doi: 10.3390/ijms222111350

Dwivedi, S. L., Heslop-Harrison, P., Spillane, C., McKeown, P. C., Edwards, D., Goldman, I., et al. (2023). Evolutionary dynamics and adaptive benefits of deleterious mutations in crop gene pools. *Trends Plant Sci.* 28, 685–697. doi: 10.1016/j.tplants.2023.01.006

EC. (2023). New techniques in biotechnology. Available online at: https://food.ec.europa.eu/plants/genetically-modified-organisms/new-techniques-biotechnology_en (Accessed 19 October 2024).

Endelman, J. B., Kante, M., Lindqvist-Kreuze, H., Kilian, A., Shannon, L. M., Caraza-Harter, M. V., et al. (2024). Targeted genotyping-by-sequencing of potato and data analysis with R/polyBreedR. *Plant Gen.* 17, e20484. doi: 10.1002/tpg2.20484

FAO. (2021). Food systems assessment – working towards the SDGs, FAO investment centre. Available online at: www.Fao.org/support-to-investment/our-ourk/projects/fsa2021/en (Accessed 17 November, 2024).

FAO. (2023). OVERVIEW - THE STATUS OF WOMEN IN AGRIFOOD SYSTEMS. Available online at: https://openknowledge.fao.org/server/api/core/bitstreams/317db554-c763-4654-a0d3-24a8488bbc3a/content/status-women-agrifood-systems-2023/overview.htmltop (Accessed 17 November, 2024).

FAO, IFAD, UNICEF, WFP and WHO. (2019). *The State of Food Security and Nutrition in the World 2019. Safeguarding against economic slowdowns and downturns* (Rome: FAO). Licence: CC BY-NC-SA 3.0 IGO.

FAODATA. (2022). Available online at: https://www.fao.org/faostat/en/homev (Accessed 17 November, 2024).

Feyissa, B. A., Amyot, L., Nasrollahi, V., Papadopoulos, Y., Kohalmi, S. E., and Hannoufa, A. (2021). Involvement of the miR156/SPL module in flooding response in medicago sativa. *Sci. Rep.* 11, 3243. doi: 10.1038/s41598-021-82450-7

Folkes, L., Moxon, S., Woolfenden, H. C., Stocks, M. B., Szittya, G., Dalmay, T., et al. (2012). PAREsnip: a tool for rapid genome-wide discovery of small RNA/target interactions evidenced through degradome sequencing. *Nucleic Acids Res.* 40, e103. doi: 10.1093/nar/gks277

Franco-Zorrilla, J. M., Valli, A., Todesco, M., Mateos, I., Puga, M. I., Rubio-Somoza, I., et al. (2007). Target mimicry provides a new mechanism for regulation of microRNA activity. *Nat. Genet.* 39, 1033–1037. doi: 10.1038/ng2079

Gahlaut, V., Zinta, G., Jaiswal, V., and Kumar, S. (2020). Quantitative epigenetics: a new avenue for crop improvement. *Epigenomes* 4, 25. doi: 10.3390/epigenomes4040025

Gaj, T., Sirk, S. J., Shui, S. L., and Liu, J. (2016). Genome-editing technologies: Principles and applications. *Cold Spring Harb Perspect. Biol.* 8 (12), a023754. doi: 10.1101/cshperspect.a023754

Gao, C. (2021). Genome engineering for crop improvement and future agriculture. *Cell* 184, 1621–1635. doi: 10.1016/j.cell.2021.01.005

Gao, Q., Luo, H., Li, Y., Liu, Z., and Kang, C. (2020). Genetic modulation of RAP alters fruit coloration in both wild and cultivated strawberry. *Plant Biotechnol. J.* 18, 1550–1561. doi: 10.1111/pbi.13317

Garcia, M. D., Nouwens, A., Lonhienne, T. G., and Guddat, L. W. (2017). Comprehensive understanding of acetohydroxyacid synthase inhibition by different herbicide families. *Proc. Natl. Acad. Sci. U.S.A.* 114, E1091–E1100. doi: 10.1073/pnas.1616142114

Garcia-Oliveira, A. L., Ayele, L., Uwimana, B., Storr, S., Taye, T., Ng, E. H., et al. (2021). *Access to genetic analysis accelerates banana breeding in East Africa*. (CGIAR: CIMMYT). Available online at: https://excellenceinbreeding.org/news/access-genetic-analysis-accelerates-banana-breeding-east-africa.

Garcia-Oliveira, A. L., Miguel, M., Acacio, J., Mutaliano, J., Abade, H., Chauque, P., et al. (2022). *Advanced genotyping in Mozambican breeding programs to accelerate genetic improvement, conservation, and agronomy*. (CGIAR: CIMMYT). Available online at: https://www.bing.com/search?q=Advanced+genotyping+in+Mozambican+breeding+programs+to+accelerate+genetic+improvement%2C+conservation%2C+and+agronomy&form=ANNH01&refig=5b80812031b9454ab2cd8828b4efdefd&pc=LCTS.

Garcia-Oliveira, A. L., Federico, M. L., Guimarães, C., Adu, G. B., Simões, F., Soares, C., et al. (2024). *The reshaping of crop breeding programs: How CGIAR genotyping and global partnerships are bridging gaps*. (CGIAR). Available online at: https://www.cgiar.org/news-events/news/the-reshaping-of-crop-breeding-programs-how-cgiar-genotyping-and-global-partnerships-are-bridging-gaps/.

Garcia-Oliveira, A. L., Martins-Lopes, P., Tolrà, R., Poschenrieder, C., Guedes-Pinto, H., and Benito, C. (2016). Differential Physiological Responses of Portuguese Bread Wheat (*Triticum aestivum* L.) Genotypes under Aluminium Stress. *Diversity* 8, 26. doi: 10.3390/d8040026

Garcia-Oliveira, A. L., Martins-Lopes, P., Tolrá, R., Poschenrieder, C., Tarquis, M., Guedes-Pinto, H., et al. (2014). Molecular characterization of the citrate transporter gene TaMATE1 and expression analysis of upstream genes involved in organic acid transport under Al stress in bread wheat (*Triticum aestivum*). *Physiol. Plant* 152, 441–452. doi: 10.1111/ppl.12179

Garcia-Oliveira, A. L., Miguel, M., Acacio, J., Mutaliano, J., et al. (2022b). *Advanced genotyping in Mozambican breeding programs to accelerate genetic improvement, conservation, and agronomy* (CGIAR). Available at: https://www.bing.com/search?q=Advanced+genotyping+in+Mozambican+breeding+programs+to+accelerate+genetic+improvement%2C+conservation%2C+and+agronomy&form=ANNH01&refig=5b80812031b9454ab2cd8828b4efdefd&pc=LCTS.

Garcia-Oliveira, A. L., Tan, L., Fu, Y., and Sun, Q. (2009). Genetic identification of quantitative trait loci for contents of mineral nutrients in rice grain. *J. Integr. Plant Biol.* 51, 84–92. doi: 10.1111/j.1744-7909.2008.00730.x

GBD 2015 Obesity Collaborators, Afshin, A., Forouzanfar, M.H., Reitsma, M.B., Sur, P., Estep, K., et al. (2017). Health effects of overweight and obesity in 195 countries over 25 years. *NEJM* 377, 13–27. doi: 10.1056/NEJMoa1614362

Ginkel, M.V., and Cherfas, J. (2023). What is wrong with biofortification. *Glob. Food Secur.* 37, 100689. doi: 10.1016/j.gfs.2023.100689

Gutierrez, A. P., Ponti, L., Kranthi, K. R., Baumgartner, J., Kenmore, P., Gilioli, G., et al. (2020). Bio-economics of Indian hybrid Bt cotton and farmer suicides. *Environ. Sci. Eur.* 32, 1–15. doi: 10.1186/s12302-020-00445-z

Hendelman, A., Stav, R., Zemach, H., and Arazi, T. (2013). The tomato NAC transcription factor SlNAM2 is involved in flower-boundary morphogenesis. *J. Exp. Bot.* 64, 5497–5507. doi: 10.1093/jxb/ert324

Hirano, T., Kazama, Y., Ishii, K., Ohbu, S., Shirakawa, Y., and Abe, T. (2015). Comprehensive identification of mutations induced by heavy-ion beam irradiation in *Arabidopsis thaliana*. *Plant J.* 82, 93–104. doi: 10.1111/tpj.12793

Hooghvorst, I., and Nogués, S. (2020). Chromosome doubling methods in doubled haploid and haploid inducer-mediated genome-editing systems in major crops. *Plant Cell Rep.* 40, 255–270. doi: 10.1007/s00299-020-02605-0

Howie, B., Fuchsberger, C., Stephens, M., Marchini, J., and Abecasis, G. R. (2012). Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat. Genet.* 44, 8, 955–8, 959. doi: 10.1038/ng.2354

ISAAA (2020). International service for the acquisition of agri-biotech applications. Bt-Cotton Increased Farmers' produce by 10-Fold in China. Available online at: https://www.isaaa.org/kc/cropbiotechupdate/article/default.asp?ID=1844 (Accessed April 23, 2024).

Ishii, T., and Araki, M. (2017). A future scenario of the global regulatory landscape regarding genome-edited crops. *GM Crops Food* 8, 44–56. doi: 10.1080/21645698.2016.1261787

Jaganathan, D., Ramasamy, K., Sellamuthu, G., Jayabalan, S., and Venkataraman, G. (2018). CRISPR for crop improvement: an update review. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.00985

Jägermeyr, J., Müller, C., Ruane, A. C., Elliot, E., Balkovic, J., Castillo, O., et al. (2021). Climate impacts on global agriculture emerge earlier in new generation of climate and crop models. *Nat. Food* 2 (11), 873–885. doi: 10.1038/s43016-021-00400-y

Jeong, I. S., Aksoy, E., Fukudome, A., Akhter, S., Hiraguri, A., Fukuhara, T., et al. (2013). Arabidopsis C-terminal domain phosphatase-like 1 functions in miRNA accumulation and DNA methylation. *PloS One* 8, e74739. doi: 10.1371/journal.pone.0074739

Jin, S., Fei, H., Zhu, Z., Luo, Y., Liu, J., Gao, S., et al. (2020a). Rationally designed APOBEC3B Cytosine Base editors with improved specificity. *Mol. Cell.* 79, 728–740.e726. doi: 10.1016/j.molcel.2020.07.005

Jin, S., Lin, Q., Luo, Y., Zhu, Z., Liu, G., Chen, K., et al. (2021). Genome-wide specificity of prime editors in plants. *Nat. Biotechnol.* 39, 1292–1299. doi: 10.1038/s41587-021-00891-x

Jørgensen, J. H. (1992). Discovery, characterization and exploitation of Mlo powdery mildew resistance in barley. *Euphytica* 63, 141–152. doi: 10.1007/BF00023919

Jouanin, A., Schaart, J. G., Boyd, L. A., Cockram, J., Leigh, F. J., Bates, R., et al. (2019). Outlook for coeliac disease patients: towards bread wheat with hypoimmunogenic gluten by gene editing of alpha- and gammagliadin gene families. *BMC Plant Biol.* 19, 333. doi: 10.1186/s12870-019-1889-5

Kakrana, A., Hammond, R., Patel, P., Nakano, M., and Meyers, B. C. (2014). sPARTA: a parallelized pipeline for integrated analysis of plant miRNA and cleaved mRNA data sets, including new miRNA target-identification software. *Nucleic Acids Res.* 42, e139. doi: 10.1093/nar/gku693

Kante, M., Lindqvist-Kreuze, H., Portal, L., David, M., and Gastelo, M. (2021). Kompetitive allele specific PCR (KASP) markers for potato: an effective tool for increased genetic gains. *Agron.* 11, 2315. doi: 10.3390/agronomy11112315

Kato, M., Masamura, N., Shono, J., Okamoto, D., Abe, T., and Imai, S. (2016). Production and characterization of tearless and non-pungent onion. *Sci. Rep.* 6, 23779. doi: 10.1038/srep23779

Kaur, N., Alok, A., Shivani, K. P., Kaur, N., Awasthi, P., Chaturvedi, S., et al. (2020). CRISPR/Cas9 directed editing of lycopene epsilon-cyclase modulates metabolic fux for betacarotene biosynthesis in banana fruit. *Metab. Eng.* 59, 76–86. doi: 10.1016/j.ymben.2020.01.008

Kelliher, T., Starr, D., Su, X., Tang, G., Chen, Z., Carter, J., et al. (2019). One-step genome editing of elite crop germplasm during haploid induction. *Nat. Biotechnol.* 37, 287–292. doi: 10.1038/s41587-019-0038-x

Khanday, I., Skinner, D., Yang, B., Mercier, R., and Sundaresan, V. (2019). A male-expressed rice embryogenic trigger redirected for asexual propagation through seeds. *Nature* 565, 91–95. doi: 10.1038/s41586-018-0785-8

Khassanova, G., Khalbayeva, S., Serikbay, D., Mazkirat, S., Bulatova, K., Utebayev, M., et al. (2023). "SNP genotyping with amplifluor-like method," in *Plant genotyping: methods and protocols, methods in molecular biology*. ed. Y. Shavrukov (New York, NY: Humana), 2638. doi: 10.1007/978-1-0716-3024-2_14

Kieu, N. P., Lenman, M., Wang, E. S., Petersen, B. L., and Andreasson, E. (2021). Mutations introduced in susceptibility genes through CRISPR/Cas9 genome editing confer increased late blight resistance in potatoes. *Sci. Rep.* 11, 4487. doi: 10.1038/s41598-021-83972-w

Kim, Y. G., Cha, J., and Chandrasegaran, S. (1996). Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc. Natl. Acad. Sci. U.S.A.* 93, 1156–1160. doi: 10.1073/pnas.93.3.1156

Knudsen, S., Wendt, T., Dockter, C., Thomsen, H. C., Rasmussen, M., Jørgensen, M. E, et al. (2021). FIND-IT: Ultrafast mining of genome diversity. *bioRxiv*. 2021.05.20.444969, 1–22. doi: 10.1101/2021.05.20.444969

Kuang, Y., Li, S., Ren, B., Yan, F., Spetz, C., Li, X., et al. (2020). Base-editing-mediated artificial evolution of osALS1 in planta to develop novel herbicide-tolerant rice germplasms. *Mol. Plant* 13, 565–572. doi: 10.1016/j.molp.2020.01.010

Kumar, K., Gambhir, G., Dass, A., Tripathi, A. K., Singh, A., Jha, A. K., et al. (2020). Genetically modified crops: current status and future prospects. *Planta* 251, 91. doi: 10.1007/s00425-020-03372-8

Kumar, D., Venkadesan, S., Prabha, R., Begam, S., Dutta, B., Mishra, D. C., et al. (2024). RiceMetaSys: Drought-miR, a one-stop solution for drought responsive miRNAs-mRNA module in rice. *Database* baae076, 1–12. doi: 10.1093/database/baae076

Kuppu, S., Ron, M., Marimuthu, M. P. A., Li, G., Huddleson, A., Siddeek, M. H., et al. (2020). A variety of changes, including CRISPR/Cas9-mediated deletions, in CENH3 lead to haploid induction on outcrossing. *Plant Biotechnol. J.* 18, 2068–2080. doi: 10.1111/pbi.13365

Lacchini, E., Kiegle, E., Castellani, M., Adam, H., Jouannic, S., Gregis, V., et al. (2020). CRISPR-mediated accelerated domestication of African rice landraces. *PloS One* 15, e0229782. doi: 10.1371/journal.pone.0229782

Laforest, L. C., and Nadakuduti, S. S. (2022). Advances in delivery mechanisms of CRISPR gene editing reagents in plants. *Front. Genome Ed.* 4. doi: 10.3389/fgeed.2022.830178

Lazzarotto, C. R., Malinin, N. L., Li, Y., Zhang, R., Yang, Y., Lee, G., et al. (2020). CHANGE-seq reveals genetic and epigenetic effects on CRISPR-Cas9 genome-wide activity. *Nat. Biotechnol.* 38, 1317–1327. doi: 10.1038/s41587-020-0555-7

Lee, R. C., Feinbaum, R. L., and Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* 75, 843–854. doi: 10.1016/0092-8674(93)90529-Y

Li, J., Jiao, G., Sun, Y., Chen, J., Zhong, Y., Yan, L., et al. (2020b). Modification of starch composition, structure and properties through editing of TaSBEIIa in both winter and spring wheat varieties by CRISPR/Cas9. *Plant Biotechnol. J.* 19, 937–951. doi: 10.1111/pbi.13519

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, T., Liu, B., Spalding, M. H., Weeks, D. P., and Yang, B. (2012). High-efficiency TALEN-based gene editing produces disease-resistant rice. *Nat. Biotechnol.* 30, 390–392. doi: 10.1038/nbt.2199

Li, J. F., Norville, J. E., Aach, J., McCormack, M., Zhang, D., Bush, J., et al. (2013). Multiplex and homologous recombination-mediated genome editing in Arabidopsis and Nicotiana benthamiana using guide RNA and Cas9. *Nat. Biotech.* 31, 688–691. doi: 10.1038/nbt.2654

Li, J., Wang, Z., He, G., Ma, L., and Deng, X. W. (2020c). CRISPR/Cas9-mediated disruption of TaNP1 genes results in complete male sterility in bread wheat. *J. Genet. Genomics* 47, 263–272. doi: 10.1016/j.jgg.2020.05.004

Li, X., Wang, Y., Sha, C., Tian, H., Daqi, F., Benzhong, Z., et al. (2018). Lycopene is enriched in tomato fruit by CRISPR/cas9-mediated multiplex genome editing. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.00559

Li, L., Wu, L. P., and Chandrasegaran, S. (1992). Functional domains in FokI restriction endonuclease. *Proc. Natl. Acad. Sci. U.S.A.* 89, 4275–4279. doi: 10.1073/pnas.89.10.4275

Li, S., Zhang, C., Li, J., Yan, L., Wang, N., and Xia, L. (2021). Present and future prospects for wheat improvement through genome editing and advanced technologies. *Plant Commun.* 2, 100211. doi: 10.1016/j.xplc.2021.100211

Liang, Z., Chen, K., Li, T., Zhang, Y., Wang, Y., Zhao, Q., et al. (2017). Efficient DNA-free genome editing of bread wheat using CRISPR/Cas9 ribonucleoprotein complexes. *Nat. Commun.* 8, 14261. doi: 10.1038/ncomms14261

Liang, X., Huang, X., and Mo, M. (2009). Feasibility analysis of mulberry breeding by aerospace mutation. *Guanxy Sericult.* 46, 32–36.

Liang, Z., Zhang, K., Chen, K., and Gao, C. (2014). Targeted mutagenesis in Zea mays using TALENs and the CRISPR/Cas system. *J. Genet. Genom* 41, 63–68. doi: 10.1016/j.jgg.2013.12.001

Liu, Y., Li, G., Zhang, Y., and Chen, L. (2019b). Current advances on CRISPR/Cas genome editing technologies in plants. *J. South China Agric. Univ.* 40, 38–49. doi: 10.7671/j.issn.1001-411X.201905058

Liu, L., Xie, Y., Guo, H., Zhao, L., Xiong, H., and Gu, J. (2021a). "New mutation techniques for crop improvement in China," in *Mutation breeding, genetic diversity and crop adaptation to climate change*. ed S. Sivasankar (International Atomic Energy Agency), 47–52. doi: 10.1079/9781789249095.000

Liu, Z., Ma, H., Jung, S., Main, D., and Guo., L. (2020b). Developmental mechanisms of fleshy fruit diversity in Rosaceae. *Annu. Rev. Plant Biol.* 71, 547–573. doi: 10.1146/annurev-arplant-111119-021700

Liu, X., Qin, R., Li, J., Liao, S., Shan, T., Xu, R., et al. (2020). A CRISPR–Cas9-mediated domain-specific base-editing screen enables functional assessment of ACCase variants in rice. *Plant Biotechnol. J.* 18, 13348. doi: 10.1111/pbi.13348

Liu, N., Tu, L., Tang, W., Gao, W., Lindsey, K., and Zhang, X. (2014). Small RNA and degradome profiling reveals a role for miRNAs and their targets in the developing fibbers of Gossypium barbadense. *TPJ* 80, 331–344. doi: 10.1111/tpj.12636

Liu, H., Wang, K., Jia, Z., Gong, Q., Lin, Z., Du, L., et al. (2019a). Efficient induction of haploid plants in wheat by editing of TaMTL using an optimized Agrobacterium-mediated CRISPR system. *J. Exp. Bot.* 71, 1337–1349. doi: 10.1093/jxb/erz529

Liu, J., Liu, X., Zhang, S., Liang, S., Luan, W., and Ma, X. (2021b). TarDB: an online database for plant miRNA targets and miRNA-triggered phased siRNAs. *BMC Gen.* 22 (1), 348. doi: 10.1186/s12864-021-07680-5

Lizumi, T., and Ramankutty, N. (2016). Changes in yield variability of major crops for 1981–2010 explained by climate change. *Environ. Res. Lett.* 11, 34003. doi: 10.1088/1748-9326/11/3/034003

Long, L., Guo, D. D., Gao, W., Yang, W. W., Hou, L. P., Ma, X. N., et al. (2018). Optimization of CRISPR/Cas9 genome editing in cotton by improved sgRNA expression. *Plant Methods* 14, 85. doi: 10.1186/s13007-018-0353-0

Lorenzo, C. D., Debray, K., Herwegh, D., Develtere, W., Impens, L., Schaumont, D., et al. (2023). BREEDIT: a multiplex genome editing strategy to improve complex quantitative traits in maize. *Plant Cell* 35, 218–238. doi: 10.1093/plcell/koac243

Lu, Y., Tian, Y., Shen, R., Yao, Q., Wang, M., Chen, M., et al. (2020). Targeted, efficient sequence insertion and replacement in rice. *Nat. Biotechnol.* 38, 1402–1407. doi: 10.1038/s41587-020-0581-5

Lv, J., Yu, K., Wei, J., Gui, H., Liu, C., Liang, D., et al. (2020). Generation of paternal haploids in wheat by genome editing of the centromeric histone CENH3. *Nat. Biotechnol.* 38, 1397–1401. doi: 10.1038/s41587-020-0728-4

Ma, L., Kong, F., Sun, K., Wang, T., and Guo, T. (2021). From classical radiation to modern radiation: Past, Present and Future of Radiation mutation breeding. *Front. Public Health* 9. doi: 10.3389/fpubh.2021.768071

Ma, Z., Ma, L., and Zhou, J. (2023). Applications of CRISPR/Cas genome editing in economically important fruit crops: recent advances and future directions. *Mol. Hortic.* 3, 1. doi: 10.1186/s43897-023-00049-0

Ma, X., Zhang, X., Liu, H., and Li, Z. (2020). Highly efficient DNA-free plant genome editing using virally delivered CRISPR-Cas9. *Nat. Plants* 6, 773–779. doi: 10.1038/s41477-020-0704-5

Madsen, C. K., Brearley, C. A., Harholt, J., and Brinch-Pedersen, H. (2024). Optimized barley phytase gene expression by focused FIND-IT screening for mutations in cis-acting regulatory elements. *Front. Plant Sci.* 15. doi: 10.3389/fpls.2024.1372049

Maheswari, M., Murthy, A. N. G., and Shanker, A. K. (2017). "Nitrogen nutrition in crops and its importance in crop quality. Nitrogen nutrition in crops and its importance in crop quality," in *The Indian nitrogen assessment*. eds. Y. P. Abrol, T. K. Adhya, V. P. Aneja, N. Raghuram, H. Pathak, U. Kulshrestha, C. Sharma and B. Singh (Elsevier Inc.), 175–186. doi: 10.1016/B978-0-12-811836-8.00012-4

Mahmoud, L. M., Kaur, P., Stanton, D., Grosser, J. W., and Dutt, M. (2022). A cationic lipid-mediated CRISPR/Cas9 technique for the production of stable genome edited citrus plants. *Plant Methods* 18, 33. doi: 10.1186/s13007-022-00870-6

Malabarba, J., Chevreau, E., Dousset, N., Veillet, F., Moizan, J., and Vergne, E. (2020). New strategies to overcome present CRISPR/Cas9 limitations in apple and pear: Efficient Dechimerization and Base Editing. *Int. J. Mol. Sci.* 22, 319–337. doi: 10.3390/ijms22010319

Marchini, J., Howie, B., Myers, S., McVean, G., and Donnelly, P. (2007). A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat. Genet.* 39, 906–913. doi: 10.1038/ng2088

Maurya, P., Sagore, B., Jain, S., Saini, S., Ingole, A., Meena, R., et al. (2022). Mutational breeding in fruit crops: A review. *Pharma. Innov.* SP-11, 1631–1637. Available online at: https://www.thepharmajournal.com/archives/2022/vol11issue6S/PartU/S-11-6-177-726.pdf.

Maxwell, D., and Hailey, P. (2020). The politics of information and analysis in famines and extreme emergencies: findings from six case studies (Feinstein International Center). Available online at: https://fic.tufts.edu/wp-content/uploads/Politics-of-Information-and-Analysis_6-5-2020.pdf (Accessed 19 October, 2024).

McIntire, J., and Dobermann, A. (2023). The CGIAR needs a revolution. *GFSI* 38, 100712. doi: 10.1016/j.gfs.2023.100712

Ming, M., Ren, Q., Pan, C., He, Y., Zhang, Y., Liu, S., et al. (2020). CRISPR-Cas12b enables efficient plant genome engineering. *Nat. Plants* 6, 202. doi: 10.1038/s41477-020-0614-6

Mokhtar, M. M., El Allali, A., Hegazy, M. E. F., and Atia, M.A.M. (2021). PlantPathMarks (PPMdb): an interactive hub for pathways-based markers in plant genomes. *Sci. Rep.* 11, 21300. doi: 10.1038/s41598-021-00504-2

Money, D., Gardner, K., Migicovsky, Z., Schwaninger, H., Zhing, G.-Y., and Myles, S. (2015). LinkImpute: fast and accurate genotype imputation for nonmodel organisms. *G3-GENES GENOM Genet.* 5, 2383–2390. doi: 10.1534/g3.115.021667

Monroe, J. G., Srikant, T., Carbonell-Bejerano, P., Becker, B., Lensink, M., Exposito-Alonso, M., et al. (2022). Mutation bias reflects natural selection in Arabidopsis thaliana. *Nature* 602, 101–105. doi: 10.1038/s41586-021-04269-6

Morin, R. D., O'Connor, M. D., Griffith, M., Kuchenbauer, F., Delaney, F. A., Prabhu, A.-L., et al. (2008). Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. *Genome Res.* 18, 610–621. doi: 10.1101/gr.7179508

Muller, H. J. (1928). The measurement of gene mutation rate in Drosophila, its high variability, and its dependence upon temperature. *Genetics* 13, 279–357. doi: 10.1093/genetics/13.4.279

Nekrasov, V., Staskawicz, B., Weigel, D., Jones, J. D., and Kamoun, S. (2013). Targeted mutagenesis in the model plant Nicotiana benthamiana using Cas9 RNA-guided endonuclease. *Nat. Biotech.* 31, 691–693. doi: 10.1038/nbt.2655

Nelson, J. W., Randolph, P. B., Shen, S. P., Everette, K. A., Chen, P. J., Anzalone, A. V., et al. (2022). Engineered pegRNAs improve prime editing efficiency. *Nat. Biotechnol.* 40, 402–410. doi: 10.1038/s41587-021-01039-7

Nielen, S., Forster, B. P., and Heslop-Harrison, J. S. (2018). "Mutagen effects in the first generation after seed treatment: biological effects of mutation treatments," in *Manual on mutation breeding, 3rd ed.* Eds. M. M. Spencer-Lopes, B. P. Forster and L. Jankuloski (Food and Agriculture Organization of the United Nations, Rome, Italy), 301.

Nishio, S., Shirasawa, K., Nishimura, R., Takeuchi, Y., Imai, A., Mase, N., et al. (2024). A self-compatible pear mutant derived from g-irradiated pollen carries an 11-Mb duplication in chromosome 17. *Front. Plant Sci.* 15. doi: 10.3389/fpls.2024.1360185

Nyika, J., and Dinka, M. O. (2023). *Water challenges in rural and urban sub-saharan africa and their management* (SpringerBriefs in Water Science and Technology). doi: 10.1007/978-3-031-26271-5_4

Okada, A., Arndell, T., Borisjuk, N., Sharma, N., Watson-Haigh, N. S., Tucker, E. J., et al. (2019). CRISPR/Cas9-mediated knockout of Ms1 enables the rapid generation of male-sterile hexaploid wheat lines for use in hybrid seed production. *Plant Biotechnol. J.* 17, 1905–1913. doi: 10.1111/pbi.13106

Oliva, R., Ji, C., Atienza-Grande, G., Huguet-Tapia, J. C., Perez-Quintero, A., Li, T., et al. (2019). Broad-spectrum resistance to bacterial blight in rice using genome editing. *Nat. Biotech.* 37, 1344–1350. doi: 10.1038/s41587-019-0267-z

Osakabe, Y., and Osakabe, K. (2015). Genome editing with engineered nucleases in plants. *Plant Cell Physiol.* 56, 389–400. doi: 10.1093/pcp/pcu170

Ouyang, X., Hong, X., Zhao, X., Zhang, W., He, X., Ma, W., et al. (2016). Knock out of the PHOSPHATE 2 gene TaPHO2-A1 improves phosphorus uptake and grain yield under low phosphorus conditions in common wheat. *Sci. Rep.* 6, 29850. doi: 10.1038/srep29850

Pan, W., Cheng, Z., Han, Z., Yang, H., Zhang, W., and Zhang, H. (2022b). Efficient genetic transformation and CRISPR/Cas9-mediated genome editing of watermelon assisted by genes encoding developmental regulators. *J. Zhejiang Univ. Sci. B.* 23, 339–344. doi: 10.1631/jzus.B2200119

Pan, C., Li, G., Malzahn, A. A., Cheng, Y., Leyson, B., Sretenovic, S., et al. (2022a). Boosting plant genome editing with a versatile CRISPR-Combo system. *Nat. Plants* 8, 513–525. doi: 10.1038/s41477-022-01151-9

Pan, C., Wu, X., Markel, K., Malzahn, A. A., Kundagrami, N., Sretenovic, S., et al. (2021). CRISPR-Act3.0 for highly efficient multiplexed gene activation in plants. *Nat. Plants.* 7, 942–953. doi: 10.1038/s41477-021-00953-7

Pastwa, E., Neumann, R. D., Mezhevaya, K., and Winters, T. A. (2003). Repair of radiation-induced DNA double-strand breaks is dependent upon radiation quality and the structural complexity of double-strand breaks. *Radiat. Res.* 159, 251–261. doi: 10.1667/0033-7587(2003)159[0251:roridd]2.0.co;2

Peng, T., Teotia, S., Tang, G. A.-O., and Zhao, Q. (2019). MicroRNAs meet with quantitative trait loci: small powerful players in regulating quantitative yield traits in rice. *WIREs RNA* 10, 1556. doi: 10.1002/wrna.1556

S. Penna and S. M. Jain (Eds.) (2023). *Mutation breeding for sustainable food production and climate resilience* (Singapure: Springer Nature), 809. doi: 10.1007/978-981-16-9720-3

Pi, M., Hu, S., Cheng, L., Zhong, R., Cai, Z., Liu, Z., et al. (2021). The MADS-box gene FveSEP3 plays essential roles in fower organogenesis and fruit development in woodland strawberry. *Hortic. Res.* 8, 247. doi: 10.1038/s41438-021-00673-1

Pompili, V., Dalla Costa, L., Piazza, S., Pindo, M., and Malnoy, M. (2020). Reduced fre blight susceptibility in apple cultivars using a high-efficiency CRISPR/Cas9- FLP/FRT-based gene editing system. *Plant Biotechnol. J.* 18, 845–858. doi: 10.1111/pbi.13253

Pray, C. E., Nagarajan, L., Huang, J., Hu, R., and Ramaswami, B. (2011). "The impact of bt cotton and the potential impact of biotechnology on other crops in China and India," in *Genetically modified food and global welfare (Frontiers of economics and globalization, vol. 10)*. Eds. C. A. Carter, G. Moschini and I. Sheldon (Emerald Group Publishing Limited, Bingley), 83–114. doi: 10.1108/S1574-8715(2011)0000010009

Puchta, H., Dujon, B., and Hohn, B. (1993). Homologous recombination in plant cells is enhanced by *in vivo* induction of double-strand breaks into DNA by a site-specific endonuclease. *Nucleic Acids Res.* 21, 5034–5040. doi: 10.1093/nar/21.22.5034

Purugganan, M. D., and Jackson, S. A. (2021). Advancing crop genomics from lab to field. *Nat. Genet.* 53, 595–601. doi: 10.1038/s41588-021-00866-3

Quiroz, D., Lensink, M., Kliebenstein, D. J., and Monroe, J. G. (2023). Causes of mutation rate variability in plant genomes. *Annu. Rev. Plant Biol.* 74, 751–775. doi: 10.1146/annurev-arplant-070522-054109

Raimondi, F., Siow, K. M., Wrona, D., Fuster-Garcia, C., Pastukhov, O., Schmitz, M., et al. (2024). Gene editing of NCF1 loci is associated with homologous recombination and chromosomal rearrangements. *Commun. Biol.* 7, 1291. doi: 10.1038/s42003-024-06959-z

Raza, A., Charagh, S., Karikari, B., Sharif, R., Yadav, V., Mubarik, M. S., et al. (2023). miRNAs for crop improvement. *Plant Physiol. Biochem.* 201, 107857. doi: 10.1016/j.plaphy.2023.107857

Ren, C., Liu, Y., Guo, Y., Duan, W., Fan, P., Li, S., et al. (2021). Optimizing the CRISPR/Cas9 system for genome editing in grape by using grape promoters. *Hortic. Res.* 8. doi: 10.3389/fpls.2019.00612

Ren, F., Ren, C., Zhang, Z., Duan, W., Lecourieux, D., Li, S., et al. (2019). Efficiency optimization of CRISPR/Cas9-mediated targeted mutagenesis in grape. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.00612

Roberts, M. G., Dawe, D., Falcon, W. P., and Naylor, R. L. (2009). El Niño-Southern Oscillation impacts on rice production in Luzon, the Philippines. *J. Appl. Meteorol. Climatol.* 48, 1718–1724. doi: 10.1175/2008JAMC1628.1

Rodriguez-Leal, D., Lemmon, Z. H., Man, J., Bartlett, M. E., and Lippman, Z. B. (2017). Engineering quantitative trait variation for crop improvement by genome editing. *Cell* 171, 470–480. doi: 10.1016/j.cell.2017.08.030

Sahu, P. K., Sao, R., Mondal, S., Vishwakarma, G., Gupta, S. K., Kumar, V., et al. (2020). Next generation sequencing based forward genetic approaches for identification and mapping of causal mutations in crop plants: A comprehensive review. *Plants* 9, 1355. doi: 10.3390/plants9101355

Sallam, A. H., Conley, E., Prakapenka, D., Da, Y., and Anderson, J. A. (2020). Improving prediction accuracy using multi-allelic haplotype prediction and training population optimization in wheat. *G3 (Bethesda)* 10, 2265–2273. doi: 10.1534/g3.120.401165

Sánchez-León, S., Gil-Humanes, J., Ozuna, C. V., Giménez, M. J., Sousa, C., Voytas, D. F., et al. (2018). Low-gluten, nontransgenic wheat engineered with CRISPR/Cas9. *Plant Biotechnol. J.* 16, 902–910. doi: 10.1111/pbi.12837

Sanei, M., and Chen, X. (2015). Mechanisms of microRNA turnover. *Curr. Opin. Plant Biol.* 27, 199–206. doi: 10.1016/j.pbi.2015.07.008

Shan, Q., Wang, Y., Li, J., and Gao, C. (2014). Genome editing in rice and wheat using the CRISPR/Cas system. *Nat. Protoc.* 9, 2395–2410. doi: 10.1038/nprot.2014.157

Shan, Q., Wang, Y., Li, J., Zhang, Y., Chen, K., Liang, Z., et al. (2013). Targeted genome modification of crop plants using a CRISPR-Cas system. *Nat. Biotechnol.* 31, 686–688. doi: 10.1038/nbt.2650

Shao, X., Wu, S., Dou, T., Zhu, H., Hu, C., Huo, H., et al. (2020). Using CRISPR/Cas9 genome editing system to create MaGA20ox2 gene-modified semi-dwarf banana. *Plant Biotechnol. J.* 18, 17–19. doi: 10.1111/pbi.13216

Shelton, A. M., Sarwer, S. H., Hossain, M., Brookes, G., and Paranjape, V. (2020). Impact of bt brinjal cultivation in the market value chain in five districts of Bangladesh. *Front. Bioeng. Biotechnol.* 8. doi: 10.3389/fbioe.2020.00498

Sheoran, S., Kumar, S., Ramtekey, V., Kar, P., Meena, R. S., and Jangir, C. K. (2022). Current status and potential of biofortification to enhance crop nutritional quality: and overview. *Sustainability* 14, 3301. doi: 10.3390/su14063301

Singh, J., Sharma, D., Brar, G. S., Sandhu, K. S., Wani, S. H., Kashyap, R., et al. (2022). CRISPR/Cas tool designs for multiplex genome editing and its applications in developing biotic and abiotic stress-resistant crop plants. *Mol. Biol. Rep.* 49, 11443–11467. doi: 10.1007/s11033-022-07741-2

Stadler, L. J. (1928a). Mutations in barley induced by x-rays and radium. *Science* 69, 186–187. doi: 10.1126/science.68.1756.186

Stadler, L. J. (1928b). Genetics effects of x-rays in maize. *PNAS* 14, 69–75. doi: 10.1073/pnas.14.1.69

Ståhlberg, A., Krzyzanowski, P. M., Egyud, M., Filges, S., Stein, L., and Godfrey, T. E. (2017). Simple multiplexed PCR-based barcoding of DNA for ultrasensitive mutation detection by next-generation sequencing. *Nat. Protoc.* 12, 664–682. doi: 10.1038/nprot.2017.006

Stoddard, B. L. (2014). Homing endonucleases from mobile group I introns: discovery to genome engineering. *Mob. DNA* 5, 1–16. doi: 10.1186/1759-8753-5-7

Stuecker, M. F., Tigchelaar, M., and Kantar, M. B. (2018). Climate variability impacts on rice production in the Philippines. *PloS One* 13, e0201426. doi: 10.1371/journal.pone.0201426

Su, Z., Bernardo, A., Tian, B., Chen, H., Wang, S., Ma, H., et al. (2019). A deletion mutation in TaHRC confers Fhb1 resistance to Fusarium head blight in wheat. *Nat. Genet.* 51, 1099–1105. doi: 10.1038/s41588-019-0425-8

Subramanian, A., and Qaim, M. (2010). The impact of bt cotton on poor households in rural India. *J. Dev. Stud.* 46, 2. doi: 10.1080/00220380903002954

Suprasanna, P., Mirajkar, S. J., and Bhagwat, S. G. (2015). "Induced mutations and crop improvement," in *Plant biology and biotechnology*. Eds. B. Bahadur, M. Venkat Rajam, L. Sahijram and K. Krishnamurthy (Springer, New Delhi). doi: 10.1007/978-81-322-2286-6_23

Svitashev, S., Schwartz, C., Lenderts, B., Young, J. K., and Cigan, A. M. (2015). Cas9-gRNA directed genome editing in maize. *Plant Physiol.* se169, 931–945. doi: 10.1038/ncomms13274

Tanaka, A., Shikazono, N., and Hase, Y. (2010). Studies on biological effects of ions beams on lethality, molecular nature of mutation, mutation rate, and spectrum of mutation phenotype for mutation breeding in higher plants. *J. Radiat. Res.* 51, 223–233. doi: 10.1269/jrr.09143

Tang, J., and Chu, C. (2017). MicroRNAs in crop improvement: fine-tuners for complex traits. *Nat. Plants* 3, 17077. doi: 10.1038/nplants.2017.77

Tang, G., Yan, J., Gu, Y., Qiao, M., Fan, R., Mao, Y., et al. (2012). Construction of short tandem target mimic (STTM) to block the functions of plant and animal microRNAs. *Methods* 58, 118–125. doi: 10.1016/j.ymeth.2012.10.006

Teotia, S., Singh, D., and Tang, G. (2020). "Technologies to address plant microRNA functions," in *Plant microRNAs concepts and strategies in plant sciences*. Eds. C. Miguel, T. Dalmay and I. Chaves (Springer, Cham.), 25–43. doi: 10.1007/978-3-030-35772-6_2

Teotia, S., Singh, D., Tang, X., and Tang, G. (2016). Essential RNA-based technologies and their applications in plant functional genomics. *Trends Biotechnol.* 34, 106–123. doi: 10.1016/j.tibtech.2015.12.001

Todesco, M., Balasubramanian, S., Cao, J., Ott, F., Sureshkumar, S., Schneeberger, K., et al. (2012). Natural variation in biogenesis efficiency of individual Arabidopsis thaliana microRNAs. *Curr. Biol.* 22, 166–170. doi: 10.1016/j.cub.2011.11.060

Tripathi, J. N., Ntui, V. O., Shah, T., and Tripathi, L. (2021). CRISPR/Cas9-mediated editing of DMR6 orthologue in banana (Musa spp.) confers enhanced resistance to bacterial disease. *Plant Biotechnol. J.* 19, 1291–1293. doi: 10.1111/pbi.13614

Tripathi, L., Ntui, V. O., and Tripathi, J. N. (2020). CRISPR/Cas9-based genome editing of banana for disease resistance. *Curr. Opin. Plant Biol.* 56, 118–126. doi: 10.1016/j.pbi.2020.05.003

Tsuruoka, C., Suzuki, M., Handle, M. P., Furusawa, Y., Anzai, K., and Okayasu, R. (2008). The difference in LET and ion species dependence for induction of initially measured and non-rejoined chromatin breaks in normal human fibroblasts. *Radiat. Res.* 170, 163–171. doi: 10.1667/RR1279.1

UN-DESA (2022). World population prospects 2022: summary of results. UN DESA/POP/2022/TR/NO. 3. Available online at: https://population.un.org/wpp/Download/Standard/MostUsed/ (Accessed 21 October, 2024).

UNDP (2022). Fostering agritech innovation: learnings from cultiv@te (Singapore: UNDP global centre for technology, innovation and sustainable development 2022). Available online at: file:///C:/Users/aoliveira/Downloads/UNDP-Cultiv@te-FinalReport-June2022-compressed.pdf (Accessed 21 October, 2024).

Urnov, F. D., Rebar, E. J., Holmes, M. C., Zhang, H. S., and Gregory, P. D. (2010). Genome editing with engineered zinc finger nucleases. *Nat. Rev. Gent.* 11, 636–646. doi: 10.1038/ngr2842

van Etten, J., de Sousa, K., Aguilarc, A., Barriosc, M., and Steinke, J. (2019). Crop variety management for climate adaptation supported by citizen science. *PNAS* 116, 4194–4199. doi: 10.1073/pnas.1813720116

Voinnet, O. (2009). Origin, biogenesis, and activity of plant microRNAs. *Cell* 136, 669–687. doi: 10.1016/j.cell.2009.01.046

Wang, Y., Cheng, X., Shan, Q., Zhang, Y., Liu, J., Gao, C., et al. (2014). Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. *Nat. Biotechnol.* 32, 947–951. doi: 10.1038/nbt.2969

Wang, J. Y., and Doudna, J. A. (2023). CRISPR technology: A decade of genome editing is only the beginning. *Science* 379, eadd8643. doi: 10.1126/science.add86

Wang, J., He, Z., Wang, G., Zhang, R., Duan, R., Gao, P., et al. (2022c). Efficient targeted insertion of large DNA fragments without DNA donors. *Nat. Methods* 19, 331–340. doi: 10.1038/s41592-022-01399-1

Wang, H., Li, Y., Chern, M., Zhu, Y., Zhang, L., Lu, J., et al. (2021). Suppression of rice miR168 improves yield, flowering time and immunity. *Nat. Plants* 7, 129–136. doi: 10.1016/j.molp.2021.12.013

Wang, C., Liu, Q., Shen, Y., Hua, Y., Wang, J., Lin, J., et al. (2019). Clonal seeds from hybrid rice by simultaneous genome engineering of meiosis and fertilization genes. *Nat. Biotechnol.* 37, 283–286. doi: 10.1038/s41587-018-0003-0

Wang, W., Simmonds, J., Pan, Q., Davidson, D., He, F., Battal, A., et al. (2018a). Gene editing and mutagenesis reveal inter-cultivar differences and additivity in the contribution of TaGW2 homoeologues to grain size and weight in wheat. *Theor. Appl. Genet.* 131, 2463–2475. doi: 10.1007/s00122-018-3166-7

Wang, Z., Wang, S., Li, D., Zhang, Q., Li, L., Zhong, C., et al. (2018b). Optimized paired-sgRNA/Cas9 cloning and expression cassette triggers high-efficiency multiplex genome editing in kiwifruit. *Plant Biotechnol. J.* 16, 1424–1433. doi: 10.1111/pbi.12884

Wang, C., Ye, J., Tang, W., Liu, Z., Zhu, C., Wang, M., et al. (2013). Loop nucleotide polymorphism in a putative miRNA precursor associated with seed length in rice (Oryza sativa L.). *Int. J. Biol. Sci.* 9, 578–586. doi: 10.7150/ijbs.6357

Wang, Z. P., Zhang, Z. B., Zheng, D. Y., Zhang, T. T., Li, X. L., Zhang, C., et al. (2022b). Efficient and genotype independent maize transformation using pollen transfected by DNA-coated magnetic nanoparticles. *J. Integr. Plant Biol.* 64, 1145–1156. doi: 10.1111/jipb.13263

Wightman, B., Ha, I., and Ruvkun, G. (1993). Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. *Cell* 75, 855–862. doi: 10.1016/0092-8674(93)90530-4

Wilson, C., Chen, P. J., Miao, Z., and Liu, D. R. (2020). Programmable m(6) A modifcation of cellular RNAs with a Cas13-directed methyltransferase. *Nat. Biotechnol.* 38, 1431–1440. doi: 10.1038/s41587-020-0572-6

Woo, J. W., Kim, J., Kwon, S. I., Corvalán, C., Cho, S. W., Kim, H., et al. (2015). DNA-free genome editing in plants with preassembled CRISPR-Cas9 ribonucleoproteins. *Nat. Biotech.* 33, 1162–1164. doi: 10.1038/nbt.3389

Xu, R., Liu, X., Li, J., Qin, R., and Wei, P. (2021). Identification of herbicide resistance OsACC1 mutations via in planta prime-editing-library screening in rice. *Nat. Plants* 7, 888–892. doi: 10.1038/s41477-021-00942-w

Xu, R., Yang, Y., Qin, R., Li, H., Qiu, C., Li, L., et al. (2016). Rapid improvement of grain weight via highly efficient CRISPR/Cas9-mediated multiplex genome editing in rice. *JGG* 43, 529–532. doi: 10.1016/j.jgg.2016.07.003

Yadav, A., Mathan, J., Dubey, A. K., and Singh, A. (2024). The emerging role of non-coding RNAs (ncRNAs) in plant growth, development, and stress response signaling. *Non-Coding RNA* 10, 13. doi: 10.3390/ncrna10010013

Yamatani, H., Kohzuma, K., Nakano, M., Takami, T., Kato, Y., Hayashi, Y., et al. (2018). Impairment of Lhca4, a subunit of LHCI, causes high accumulation of chlorophyll and the stay-green phenotype in rice. *J. Exp. Bot.* 69, 1027–1035. doi: 10.1093/jxb/erx468

Yang, Z., Hui, S., Lv, Y., Zhang, M., Chen, D., Tian, J., et al. (2022). miR395-regulated sulfate metabolism exploits pathogen sensitivity to sulfate to boost immunity in rice. *Mol. Plant* 15, 671–688. doi: 10.1016/j.molp.2021.12.013

Yoon, D.-K., Sugamani, M., Ishiyama, K., Kagawa, T., Tanaka, M., Nagao, R., et al. (2022). The gs3allele from a large-grain rice cultivar, Akita63, increases yield and improves nitrogen-use efficiency. *Plant Direct* 6, e417. doi: 10.1002/pld3.417

Yu, H., Lin, T., Meng, X., Du, H., Zhang, J., Liu, G., et al. (2021). A route to *de novo* domestication of wild allotetraploid rice. *Cell* 184, 1156–1170. doi: 10.1016/j.cell.2021.01.013

Zhang, Y., Bai, Y., Wu, G., Zou, S., Chen, Y., Gao, C., et al. (2017). Simultaneous modification of three homoeologs of TaEDR1 by genome editing enhances powdery mildew resistance in wheat. *Plant J.* 91, 714–724. doi: 10.1111/tpj.13599

Zhang, C., Ding, Z., Wu, K., Yang, L., Li, Y., Yang, Z., et al. (2016a). Suppression of jasmonic acid-mediated defense by viral-inducible microRNA319 facilitates virus infection in rice. *Mol. Plant* 9, 1302–1314. doi: 10.1016/j.molp.2016.06.014

Zhang, Y., Li, D., Zhang, D., Zhao, X., Cao, X., Dong, L., et al. (2018b). Analysis of the functions of TaGW2 homoeologs in wheat grain weight and protein content traits. *Plant J.* 94, 857–866. doi: 10.1111/tpj.13903

Zhang, Y., Liang, Z., Zong, Y., Wang, Y., Liu, J., Chen, K., et al. (2016). Efficient and transgene-free genome editing in wheat through transient expression of CRISPR/Cas9 DNA or RNA. *Nat. Commun.* 7, 12617. doi: 10.1038/ncomms12617

Zhang, R., Liu, J., Chai, Z., Chen, S., Bai, Y., Zong, Y., et al. (2019a). Generation of herbicide tolerance traits and a new selectable marker in wheat using base editing. *Nat. Plants* 5, 480–485. doi: 10.1038/s41477-019-0405-0

Zhang, Y., Malzahn, A. A., Sretenovic, S., and Qi, Y. (2019). The emerging and uncultivated potential of CRISPR technology in plant science. *Nat. Plants.* 5, 778–794. doi: 10.1038/s41477-019-0461-5

Zhang, B., and Unver, T. (2018). A critical and speculative review on microRNA technology in crop improvement: Current challenges and future directions. *Plant Sci.* 274, 193–200. doi: 10.1016/j.plantsci.2018.05.031

Zhang, Y., Wang, H., Du, Y., Zhang, L., Li, X., Guo, H., et al. (2024). Biological responses of an elite centipedegrass [Eremochloa ophiuroides (Munro) Hack.] cultivar (Ganbei) to carbon ion beam irradiation. *Front. Plant Sci.* 15. doi: 10.3389/fpls.2024.1433121

Zhang, S., Wu, S., Hu, C., Yang, Q., Dong, T., Sheng, O., et al. (2022). Increased mutation efficiency of CRISPR/Cas9 genome editing in banana by optimized construct. *Peer J.* 10, e12664. doi: 10.7717/peerj.12664

Zhang, J., Zhang, H., Srivastava, A. K., Pan, Y., Bai, J., Fang, J., et al. (2018a). Knockdown of rice microRNA166 confers drought resistance by causing leaf rolling and altering stem xylem development. *Plant Physiol.* 176, 2082–2094. doi: 10.1104/pp.17.01432

Zhao, H., Guo, H., Gu, J., Zhao, S., Li, J., and Liu, L. (2010). A temperature-sensitive winter wheat chlorophyll mutant derived from space mutated. *J. Nucl. Agric.* 24, 1110–1116.

Zhao, M., Liu, B., Wu, K., Ye, Y., Huang, S., Wang, S., et al. (2015). Regulation of osmiR156h through alternative polyadenylation improves grain yield in rice. *PloS One* 10, e0126154. doi: 10.1371/journal.pone.0126154

Zhaozhi, L., Xiaojie, H., XiaoXian, L., Chunhong, Y., Downes, S., Parry, H., et al. (2022). Quo vadis Bt cotton: a dead-end trap crop in the post Bt era in China? *Entomol. Gen.* 42, 649–654. doi: 10.1127/entomologia/2021/1355

Zheng, X., Tang, X., Wu, Y., Zheng, X., Zhou, J., Han, Q., et al. (2024). An efficient CRISPR-Cas12a-mediated MicroRNA knockout strategy in plants. *Plant Biotech. J.*, 1–13. doi: 10.1111/pbi.14484

Zhong, Y., Blennow, A., Kofoed-Enevoldsen, O., Jiang, D., and Hebelstrup, K. H. (2019b). Protein Targeting to Starch 1 is essential for starchy endosperm development in barley. *J. Exp. Bot.* 70, 485–496. doi: 10.1093/jxb/ery398

Zhong, Y., Liu, C., Qi, X., Jiao, Y., Wang, D., Wang, Y., et al. (2019a). Mutation of *ZmDMP* enhances haploid induction in maize. *Nat. Plant* 5, 575–580. doi: 10.1038/s41477-019-0443-7

Zhou, J., Wang, G., and Liu, Z. (2018). Efficient genome editing of wild strawberry genes, vector development and validation. *Plant Biotechnol. J.* 16, 1868–1877. doi: 10.1111/pbi.12922

Zhu, H., Li, C., and Gao, C. (2020). Applications of CRISPR-Cas in agriculture and plant biotechnology. *Nat. Rev. Mol. Cell Biol.* 21, 661–677. doi: 10.1038/s41580-020-00288-9

Glossary

AgMIP — Agricultural Model Intercomparison and Improvement Project is a collaborative global research effort aiming to enhance agricultural models's accuracy in predicting crop yields, food security assessment, and climate change impacts on agriculture and was established in 2010

CRISPR — Clustered Regularly Interspaced Short Palindromic Repeat system used to modify/edit a cultivar of plant species. It facilitates site-directed mutagenesis

CRISPR off-target — Accidental or unintentional alterations in different DNA regions, different from the intended spots, with similar but not exact sequences to the target site

Crop productivity — Efficiency of crop production, measured as yield *per* unit area. It can exceed 100% (as shown in Figure 1 of this paper), when yields surpass previous standards (global average) due to improved varieties, better management practices, or favourable conditions

DSB — Double-strand break (DSB) mutations and refers to a type of genetic alteration in which both strands of the DNA molecule are severed or broken at the same location. This disruption in the DNA structure can lead to various genetic changes, including mutations, deletions, insertions, and rearrangements

KASP™ — Kompetitive Allele-Specific PCR, is a genotyping technology used to detect single nucleotide polymorphisms (SNPs) and small insertions or deletions in DNA samples

KASP haplotyping — Technology that allows the detection of SNPs in DNA sequences. It is a form of competitive PCR where allele-specific primers are designed to selectively amplify target DNA sequences. KASP haplotyping specifically focuses on identifying and characterizing haplotypes. Haplotypes are sets of closely linked genetic markers that tend to be inherited together. Both KASP haplotyping and MAS contribute to genomic selection by providing genetic information, KASP haplotyping specifically focuses on haplotype structures, whereas MAS encompasses a broader range of marker-assisted techniques for selecting individuals based on genetic markers associated with traits

IGIs — Inclusive genomic Innovations

IPRs — Intellectual Property Rights, are legal protections granted to individuals or entities for their creations of the mind or intellect

ISIMIP — The Inter-Sectoral Impact Model Intercomparison Project is an international research initiative aimed at assessing climate change's potential impacts on diverse sectors, extending beyond agriculture to encompass water resources, ecosystems, health, economics, and more

Mutation breeding — The use of mutagens, such as radiation or chemicals, to induce genetic mutations in plants, leading to increased genetic variability generation of new allele variation for trait improvement

Mutational signature marker — These are specific patterns of mutations found in DNA indicating the underlying processes or exposures that caused those mutations. These signatures can be associated with various factors, including environmental influences and intrinsic factors, and adapted from Human genetics

NGT — New Genomic Technologies are a variety of techniques that alter the genetic material of an organism. Those considered equivalent to conventional plants (NGT 1 plants) would be exempted from most requirements of the GMO legislation. https://www.europarl.europa.eu/news/en/agenda/briefing/2024-02-05/2/new-genomic-techniques-debate-and-vote-on-new-eu-rules

NUE — Nitrogen Use Efficiency is the measure of how effectively plants utilize nitrogen from fertilizers, encompassing uptake for growth and minimizing losses through processes like leaching and volatilization

QC — Quality control ensures that the genetic data used in research is robust, reliable, and suitable for interpretation and analysis. By implementing QC measures, breeders can verify the genetic purity of the F_1 offspring, thereby confirming these are genuine hybrids derived from the intended parental lines. Moreover, QC in breeding programs supports the monitorization of the genetic diversity within the populations. It ensures that breeding stocks retain a broad spectrum of alleles, and therefore, reducing the risk of inbreeding depression and preserving the potential for future genetic improvement

STTMs — Short Tandem Target Mimic are a class of synthetic small RNAs that are designed to regulate the expression of specific genes in plants by modulating microRNA (miRNA) activity

TALENs — Transcription Activator-Like Effector Nucleases, are a type of engineered proteins used for targeted genome editing in various organisms, including plants, animals, and even human cells

TILLING — Targeting Induced Local Lesions IN Genomes, is a high-throughput method used for identifying and characterizing mutations in specific genes of interest

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 12 February 2025
DOI 10.3389/fpls.2025.1533219

OPEN ACCESS

EDITED BY
Diaa Abd El Moneim,
Arish University, Egypt

REVIEWED BY
Muhammad Saad Shoaib Khan,
Jiangsu University, China
Kamran Shah,
South China Agricultural University, China

*CORRESPONDENCE
Chunyun Guan
guancy201l@aliyun.com
Mei Guan
gm7142005@hunau.edu.cn

RECEIVED 23 November 2024
ACCEPTED 16 January 2025
PUBLISHED 12 February 2025

CITATION
Yao M, Hong B, Ji H, Guan C and Guan M (2025) Genome-wide identification of PDX and expression analysis under waterlogging stress exhibit stronger waterlogging tolerance in transgenic *Brassica napus* plants overexpressing the *BnaPDX1.3* gene compared to wild-type plants. *Front. Plant Sci.* 16:1533219. doi: 10.3389/fpls.2025.1533219

Genome-wide identification of PDX and expression analysis under waterlogging stress exhibit stronger waterlogging tolerance in transgenic *Brassica napus* plants overexpressing the *BnaPDX1.3* gene compared to wild-type plants

Mingyao Yao[1,2], Bo Hong[1,2], Hongfei Ji[1,2], Chunyun Guan[1,2,3]* and Mei Guan[1,2,3]*

[1]College of Agriculture, Hunan Agricultural University, Changsha, China, [2]Hunan Branch of National Oilseed Crops Improvement Center, Changsha, China, [3]Southern Regional Collaborative Innovation Center for Grain and Oil Crops in China, Changsha, China

The *PDX* gene is a key gene in the vitamin B6 synthesis pathway, playing a crucial role in plant growth, development, and stress tolerance. To explore the family characteristics of the *PDX* gene in *Brassica napus* (*B. napus*) and its regulatory function under waterlogging stress, this study used five *PDX* genes from Arabidopsis thaliana as the basis for sequence analysis. Thirteen, eight, and six *PDX* genes were identified in *B. napus*, *Brassica oleracea* (*B. oleracea*), and *Brassica rapa* (*B. rapa*), respectively. Bioinformatics study reveals high conservation of *PDX* subfamily genes during evolution, and *PDX* genes in *B. napus* respond to waterlogging stress.In order to further investigate the effect of the *PDX* gene on waterlogging tolerance in *B. napus*, expression analysis was conducted on *BnaPDX1.3* gene overexpressing *B. napus* plants and wild-type plants. The study showed that overexpressing plants could synthesize more VB6 under waterlogging stress, exhibit stronger antioxidant enzyme activity, and have a more effective and stable ROS scavenging system, thus exhibiting a healthier phenotype. These findings suggested that the *BnaPDX1.3* gene can enhance the waterlogging tolerance of *B. napus*, which is of great significance for its response to waterlogging stress. Our study provides a basic reference for further research on the regulation mechanism of the *PDX* gene and waterlogging tolerance in *B. napus*.

KEYWORDS

Brassica napus, BnaPDX1.3, waterlogging stress, vitamin B6, overexpression

1 Introduction

Vitamin B6 is an essential water-soluble vitamin required by all living organisms (Gorelova et al., 2022). It is recognized as an antioxidant and is linked to responses to various biotic and abiotic stresses (Vanderschuren et al., 2013; Samsatly et al., 2016; Chung, 2012; Zhang et al., 2015; Asensi-Fabado and Munné-Bosch, 2010). Vitamin B6 consists of six interconvertible compounds: pyridoxine phosphate (PNP), pyridoxal phosphate (PLP), pyridoxamine phosphate (PMP), and their nonphosphorylated derivatives (pyridoxine [PN], pyridoxal [PL], and pyridoxamine [PM]) (Fitzpatrick et al., 2007; Roje, 2007; Mooney et al., 2009). Additionally, vitamin B6 must be phosphorylated to function as a coenzyme (Colinas et al., 2016). Among them, PLP is the most important coenzyme, playing a crucial role in lipid degradation and carbohydrate storage (such as glycogen) (Jeong and VacantiLi, 2020; Mooney et al., 2009; Rueschhoff et al., 2013). It also plays a decisive role in amino acid metabolism, catalyzing transamination, decarboxylation, and α,β-elimination reactions involved in amino acid metabolism (Eliot and Kirsch, 2004; Drewke and Leistner, 2001). Other studies have found that the five enzymes that most commonly use PLP as a coenzyme, in order of dependency, are transferases, lyases, isomerases, hydrolases, and oxidoreductases, demonstrating the versatility of PLP-dependent enzymes (Percudani and Peracchi, 2003).

In plants, the *de novo* biosynthesis pathway of vitamin B6 relies on two proteins, *PDX1* and *PDX2*, which function as glutaminyltransferases (Tambasco-Studart et al., 2005, 2007). *PDX2* possesses transaminase activity, extracting ammonium groups from glutamine and incorporating them into the product. *PDX1* receives these ammonium groups and synthesizes the final product (Raschle et al., 2005; Dong et al., 2004). In the *AtPDX* family of *Arabidopsis thaliana*, there are three *AtPDX1* (*AtPDX1.1*, *AtPDX1.2*, and *AtPDX1.3*) and one *AtPDX2* protein member (Tambasco-Studart et al., 2005; Wetzel et al., 2004). However, only *PDX1.1* and *PDX1.3* are involved in the biosynthesis of vitamin B6, while *PDX1.2* does not play a role in this process (Tambasco-Studart et al., 2005; Titiz et al., 2006). Instead, *PDX1.2* functions as a pseudoenzyme that enhances the activity of catalytic homologs under stress conditions (Moccand et al., 2014). Studies on *pdx1* and *pdx2* mutants in *Arabidopsis thaliana* have shown that knocking out both genes, *PDX1.1* and *PDX1.3*, or knocking out the single gene, *PDX2*, leads to the death of the mutant at the embryonic stage of development (Tambasco-Studart et al., 2005; Titiz et al., 2006). However, single mutants of *pdx1.1* or *pdx1.3* can survive, although with a short root phenotype, with the latter phenotype being more pronounced (Titiz et al., 2006; Chen and Xiong, 2005; Wagner et al., 2006; Boycheva et al., 2015). Hao Chen et al. (2009) suggested that the short-root phenotype in *pdx1* mutants was caused by reduced levels of endogenous auxin synthesis (Chen and Xiong, 2009). In *Arabidopsis thaliana*, the expression level of *PDX1.3* is consistently higher than that of *PDX1.1*, both spatially and temporally. Additionally, *pdx1.3* mutants exhibit more pronounced phenotypic differences in morphology and development compared to *pdx1.1* mutants (Titiz et al., 2006), despite the two proteins being 87% identical and capable of synthesizing vitamin B6 at comparable rates (Tambasco-Studart et al., 2005). In addition, Leuendorf et al. (2014) demonstrated that the lack of *PDX1.2* affects seed and hypocotyl development and results in a large number of aborted seeds during embryonic development, as shown through T-DNA insertion-generated heterozygous *PDX1.2* and artificial microRNA-reduced *PDX1.2* expression.

The *PDX* gene is not only crucial for the growth and development of plants but also indispensable for their response to adverse environmental stresses. Ristilä et al. (2011) discovered that *PDX1.3* is involved in plants' response to UV-B radiation, with the formation of UV-B-induced *PDX1.3* primarily occurring in the parts of leaves that absorb UV-B radiation (Ristilä et al., 2011). It has also been proven that the *pdx1.3* mutant of *Arabidopsis thaliana* exhibits weaker tolerance to strong light and photo-oxidation (Havaux et al., 2009). Additionally, the *PDX1.2* gene of *Arabidopsis thaliana* has been found to exhibit higher expression under UV-B treatment, oxidative stress, and heat shock (Denslow et al., 2007). Furthermore, Dell'Aglio et al. (2017) identified a heat stress transcription initiation site within *PDX1.2*, and subsequent research indicated that *PDX1.2* can stably catalyze *PDX1* homologs under heat stress conditions.

Research on vitamin B6 (VB6) and *PDX* genes primarily focuses on their roles in plant growth, development, and stress resistance. However, limited studies have examined their roles in the response of *Brassica napus* (*B. napus*) to waterlogging stress. In this study, we investigated the impact of *BnaPDX1.3* gene overexpression (PDX1.3#20 and PDX1.3#21) and WT *B. napus* plants on waterlogging tolerance.

2 Materials and methods

2.1 Experimental materials

The *B. napus* variety G218 was provided by the Hunan Branch of the National Oil Crop Improvement Center.

2.2 *PDX* gene family analysis

Information related to the *PDX* gene of *Arabidopsis thaliana*, *Brassica rapa* (*B. rapa*), *Brassica oleracea* (*B. oleracea*), and *B. napus* was obtained from the Ensembl Plants database (http://plants.ensembl.org/index.html), including the full-length DNA sequence, CDS sequence, amino acid sequence, etc. The nucleic acid sequence and amino acid sequences of the *PDX* gene of *Arabidopsis thaliana* were first downloaded from Ensembl Plants. BLASTP alignment was performed on the genomes of *B. rapa*, *B. oleracea*, and *B. napus* with a threshold of $E < 10^{-5}$ to identify *PDX* homologous genes. A phylogenetic tree was then constructed using the Neighbor-Joining (NJ) method in MEGA 11 (Tamura et al., 2021). Subsequently, Evolview (http://www.evolgenius.info/evolview/) was used to enhance the visualization of the phylogenetic tree (Subramanian et al., 2019). Motif analysis was conducted on the *PDX* gene of *Arabidopsis thaliana*, *B. rapa*, *B.*

oleracea, and *B. napus* using MEME (http://MEME.nbcr.net/MEME/cgi-bin/MEME.cgi) (Bailey et al., 2015). Chromosome position information for the *PDX* gene was obtained from the gtf file of the Ensembl Plants database, and chromosome localization was analyzed using TBtools software (Chen et al., 2020). *Cis*-element analysis was conducted on the promoter region of the DNA sequence upstream 2,000 bp of the *PDX* gene of *Arabidopsis thaliana*, *B. rapa*, *B. oleracea*, and *B. napus* using PlantCARE (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) (Lescot et al., 2002). The expression pattern of the *BnaPDX* gene under waterlogging stress was analyzed based on the data from Hong et al. (2023). The results of gene structure, conserved domains, chromosome localization, *cis*-elements in the promoter region, and RNA-seq analysis were displayed using TBtools software (Chen et al., 2020).

2.3 qRT-PCR measurement

Total RNA was extracted from *B. napus* tissues using an assay kit from Vazyme Biotech Co. Ltd. (Nanjing, China) and reverse transcribed into cDNA with a reverse transcription assay kit from Takara Biomedical Technology (Beijing, China) Co. Ltd. Based on the reference gene sequence of *BnaPDX1.3* in *B. napus*, specific fluorescent quantitative primers were designed using the Primer-BLAST tool on the NCBI website (Supplementary Table S1). Following the instructions of the fluorescent quantitative PCR assay kit from TransGen Biotech Co. Ltd. (Beijing China)., the gene expression was detected using a Bio-Rad fluorescent quantitative PCR instrument. The data were analyzed using the $2^{-\Delta\Delta T}$ method, with the *Bnactin* gene as the internal reference gene, to determine the relative expression level of the gene (Supplementary Table S1) (Livak and Schmittgen, 2001).

2.4 Construction of overexpression vectors and *Brassica napus* transformation

Based on the PC2300S vector map, we selected the *Kpn*I and *Xba*I restriction sites for constructing the *35S::BnaC03.PDX1.3* overexpression vector. We amplified the target fragment containing the restriction sites (Supplementary Table S2; Supplementary Figure S1), ligated the target vector with T4 ligase after restriction digestion, and used double enzyme digestion to verify the construction status of the vector (Supplementary Figure S2). We selected the successfully constructed recombinant overexpression vector, transferred it into *B. napus* G218 via *Agrobacterium*-mediated transformation, and performed positive detection on the transformed seedlings using specific primers for the *NptII* gene (Supplementary Table S3; Supplementary Figure S3). We used designed specific primers for the *35S::BnaC03.PDX1.3* recombinant overexpression vector to perform positive detection on T1 transgenic plants (Supplementary Table S3; Supplementary Figures S4, S5) and screened for *BnaPDX1.3* gene overexpression in the T1 generation plants.

2.5 Biomass measurement

Two groups of T1 generation *B. napus* G218 plants overexpressing the *BnaPDX1.3* gene and one group of G218 wild-type plants were subjected to 15 days of waterlogging stress. After the treatment, the fresh weight of the plants from all three groups was measured, and biomass analysis was conducted.

2.6 Vitamin B6 content measurement

Leaf tissues were collected from two groups of *BnaPDX1.3*-overexpressing *B. napus* plants and one group of wild-type plants after exposure to waterlogging for 0, 2, 4, 6, 9, 12, and 15 days. These tissues were flash-frozen in liquid nitrogen and stored at − 80°C. The VB6 content of the samples was detected using a vitamin B6 assay kit provided by ZCIBIO Technology Co. Ltd. (Shanghai, China).

2.7 Antioxidant enzyme activity and H_2O_2 content measurement

Leaf tissues were collected from two groups of *BnaPDX1.3*-overexpressing *B. napus* plants and one group of wild-type plants after exposure to waterlogging for 0, 2, 4, 6, 9, 12, and 15 days, respectively. The tissues were flash-frozen in liquid nitrogen and stored at − 80°C. The antioxidant enzyme activity and H_2O_2 content of the samples were detected using the SOD, POD, CAT, and H_2O_2 assay kits provided by ZCIBIO Technology Co. Ltd.

3 Statistical analysis

All measurements were conducted using three biological replicates. One-way ANOVA followed by Dunnett's multiple comparisons test was performed using GraphPad Prism (v. 25), with $p < 0.05$ considered significant for all experiments.

4 Results

4.1 *PDX* gene family analysis

4.1.1 Phylogenetic analysis of *PDX* proteins in *B. napus*, *B. rapa*, and *B. oleracea*

Conducting a BLAST search on the protein sequences (*ATPDX*) of five *Arabidopsis thaliana PDX* genes yielded six *B. rapa* genes (*BraPDX*), eight *B. oleracea* genes (*BoPDX*), and 13 *B. napus* genes (*BnaPDX*), all obtained from the Ensembl Plants database. Based on the *PDX* protein sequences from the four species, an evolutionary

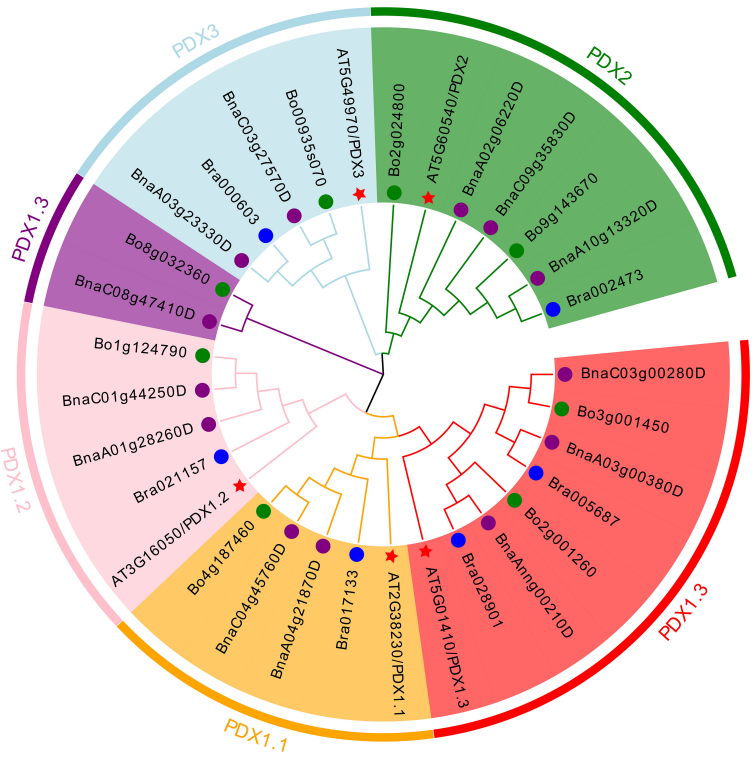


FIGURE 1
Phylogenetic tree of *PDX* proteins from *B. napus*, *B. rapa*, *B. oleracea*, and *Arabidopsis*. The neighbor-joining phylogenetic tree shows the relationships among 13 *B. napus*, six *B. rapa*, eight *B. oleracea*, and five *Arabidopsis PDX* proteins.

tree (Figure 1) was constructed using MEGA 11 software and the Evolview online beautification tool, visually reflecting the evolution and classification of the 32 *PDX* family members. According to the evolutionary tree analysis, the *PDX* proteins were divided into five subclades (*PDX1.1*, *PDX1.2*, *PDX1.3*, *PDX2*, *PDX3*), with two branches in *PDX1.3*. All five subclades contained genes from *B. rapa*, *B. oleracea*, and *B. napus*.

4.1.2 Motif analysis (MEME) and gene structural analysis of *PDX*

To further investigate the evolutionary relationships among the *PDX* gene families of *Arabidopsis thaliana*, *B. rapa*, *B. oleracea*, and *B. napus*, we conducted a predictive analysis of their gene structures and motifs. Using MEME, we analyzed 15 motifs among 32 proteins (Figure 2). It was found that members belonging to the same subfamily exhibited certain similarities in their conserved motifs. The conserved motifs of most *PDX1.1* and *PDX1.3* proteins were arranged as Motif11-Motif3-Motif6-Motif2-Motif4-Motif1-Motif5-Motif13, with the exception of *Bra017133*, which lacked Motif4, Motif6, and Motif13; *BnaA04g21870D*, which lacked Motif4 and Motif6; *BnaC04g45760D*, which lacked Motif2; and *BnaC08g47410D* and *Bo8g032360*, which retained only the core conserved motif, Motif1, of the pyridoxal phosphate synthase domain. This may also imply that they still perform functions related to vitamin B6 synthesis. Compared to *PDX1.1* and *PDX1.3*, the conserved motif of the *PDX1.2* protein lacked Motif13, with its composition pattern being Motif11-Motif3-Motif6-Motif2-Motif4-Motif1-Motif5. In summary, the functions of *PDX1.1*, *PDX1.2*, and *PDX1.3* proteins should be similar.

Except for *AT5G49970*, which lacks Motif1, Motif9, Motif12, and Motif14, the conserved motif patterns of the remaining *PDX2* proteins are Motif5-Motif11-Motif8-Motif15-Motif10-Motif1-Motif12-Motif14-Motif9. Most *PDX3* proteins have conserved motif patterns of Motif11-Motif9-Motif7-Motif14-Motif5-Motif8, with *Bo2g024800* lacking Motif8 and *BnaC09g35830D* lacking Motif9 and Motif11. The motif prediction results are generally consistent with the evolutionary tree alignment analysis, suggesting that *PDX1* genes from different subclades may perform similar functions while exhibiting significant differences from the other two major clades. This motif specificity distribution among different subclades may reflect the functional differentiation of *BnaPDX* genes in *B. napus*.

Gene structure prediction reveals that the structure of the *PDX1* genes (*PDX1.1*, *PDX1.2*, *PDX1.3*) is relatively conserved, with most introns ranging from 0 to 2 in number. Specifically, 13 *PDX1* genes lack introns, three *PDX1* genes possess one intron and two exons, three *PDX1* genes have two introns and three exons, and only *Bra017133* features nine introns and 10 exons. Meanwhile, the gene structure of the *PDX2* subgroup remains consistent, exhibiting a structure with 12 introns and 13 exons. Within the *PDX3* subgroup, two types exhibit four introns and five exons: *Bo2g024800* and *BnaA02g06220D*, while the rest possess a structure with five introns and six exons.

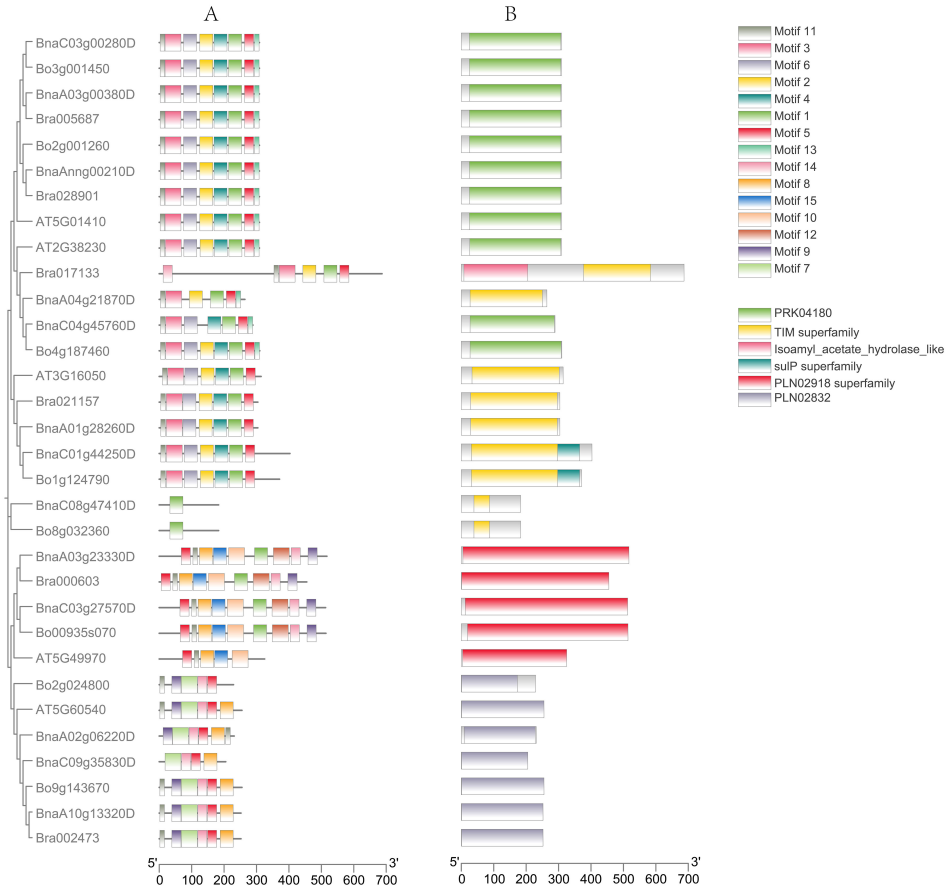


FIGURE 2
Gene motifs **(A)** and gene structure **(B)** analysis of *PDX* proteins. Fifteen motifs in *PDX* proteins were identified using MEME tools **(A)**. Orange boxes, black lines, and green boxes indicate exons, introns, and untranslated regions, respectively **(B)**.

4.1.3 Chromosomal distribution of *PDX* genes

The *PDX* gene does not exhibit tandem repeats in the four species. Five *Arabidopsis thaliana ATPDX* genes are located on chromosomes Chr2, Chr3, and Chr5; six Chinese *B. rapa BraPDX* genes are located on chromosomes A01, A02, A03, A04, and A10; and eight *B. oleracea BoPDX* genes and 13 *B. napus BnaPDX* genes are located on seven and 11 chromosomes, respectively (Figure 3).

4.1.4 The *cis*-element regulators in *BnaPDX1.3* promoters

Using PlantCARE to analyze the various *cis*-elements in the promoter sequence (upstream of 2,000 bp) of the *BnaPDX* gene, the results can be divided into five types of plant hormone response *cis*-elements (IAA, ABA, MeJA, GA, and SA), six types of environmental stress response *cis*-elements (Stress, Anaerobic, Circadian, Light, Drought, and Low-T), one specific transcription factor binding site (MYBHv1), and several gene-specific response *cis*-elements (Figure 4).

The results revealed that within the *BnaPDX* gene family of *B. napus*, there are 111 *cis*-elements linked to plant hormone response, 239 *cis*-elements associated with environmental stress response, and five *cis*-elements related to metabolism. Specifically, among these, 44 ABA response *cis*-elements represent the highest proportion within plant hormone response, whereas 169 light response *cis*-elements are the most prevalent in environmental stress response.

4.2 *BnaPDX1.3* of *B. napus* responds to waterlogging stress

To further analyze the response of *PDX* members to waterlogging stress, this study utilized RNA-seq data from *B. napus* G218 subjected to waterlogging treatment for 6 days, previously conducted in the field waterlogging experiment by our research group (Hong et al., 2023). We analyzed the expression of the *BnaPDX* gene family in *B. napus* under waterlogging stress and performed qRT-PCR verification of the *BnaPDX1.3* gene (*BnaAnng00210D*, *BnaA03g00380D*, *BnaC03g00280D*).

Among the *BnaPDX* family genes, except for *BnaPDX1.2* (*BnaA01g28260D* and *BnaC01g44250D*) whose expression level was downregulated after 6 days of waterlogging stress, all other *BnaPDX* family genes showed upregulation, with *BnaPDX1.1* (*BnaA04g21870D* and *BnaC04g45760D*) exhibiting the most significant upregulation (Figure 5A). The qRT-PCR verification results were generally consistent with the RNA-seq data. Except for the *BnaAnng00210D* gene, whose expression decreased relatively

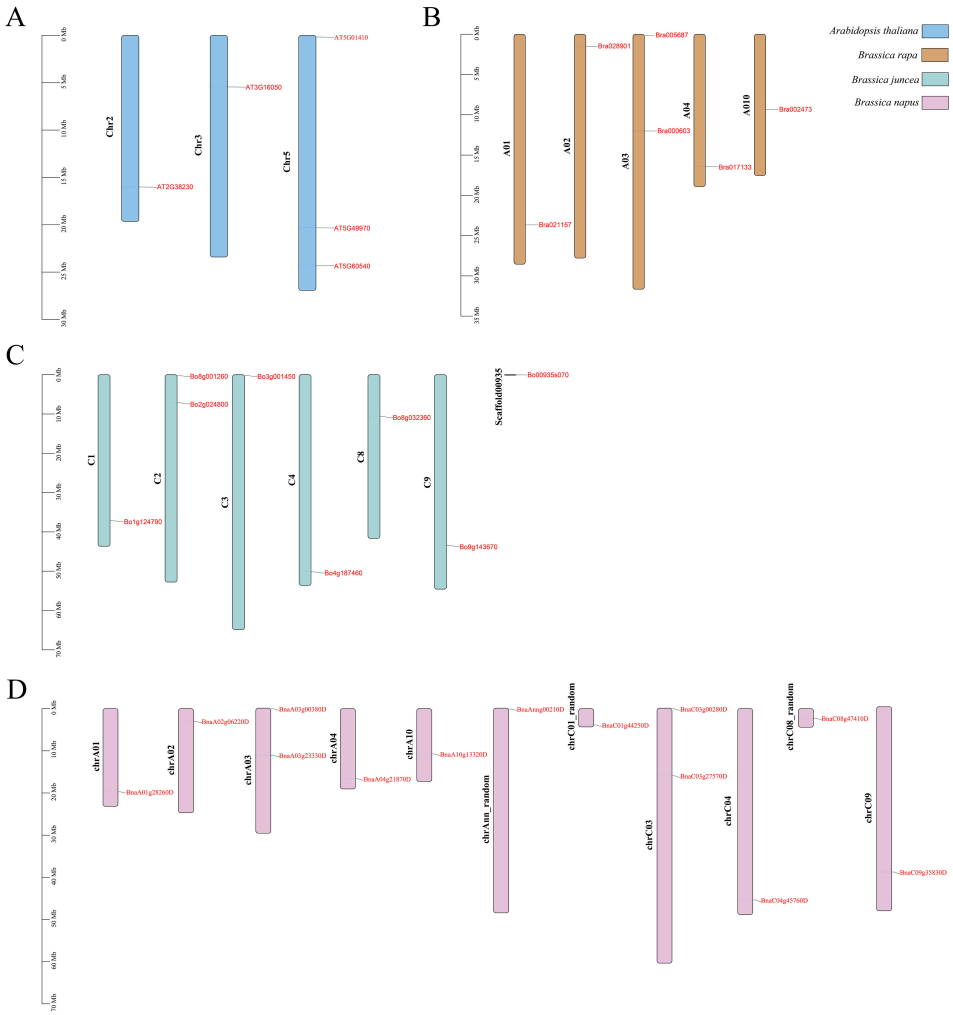


FIGURE 3
Chromosomal distribution of *PDX* genes in *Arabidopsis* **(A)**, *B. rapa* **(B)**, *B. oleracea* **(C)**, and *B. napus* **(D)**.

after 6 days of waterlogging stress, the expression of the *BnaPDX1.3* gene was upregulated after stress in all other cases.

Interestingly, *BnaAnng00210D* exhibits the highest relative expression level, being 24 times higher than that of *BnaA03g00380D*. Its elevated expression level might account for its slightly downregulated expression under waterlogging stress (Figure 5B). Further investigation into the expression pattern of *BnaPDX1.3* under waterlogging stress revealed that the relative

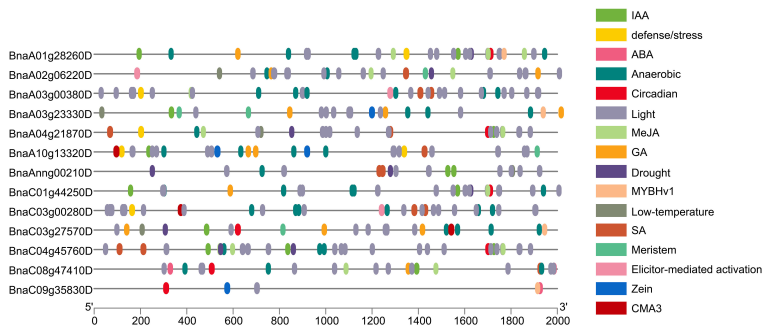


FIGURE 4
Promoter *cis*-element analysis of *PDX* genes in *B. napus*. Differently colored boxes represent various *cis*-acting elements and their corresponding responses.

10.3389/fpls.2025.1533219

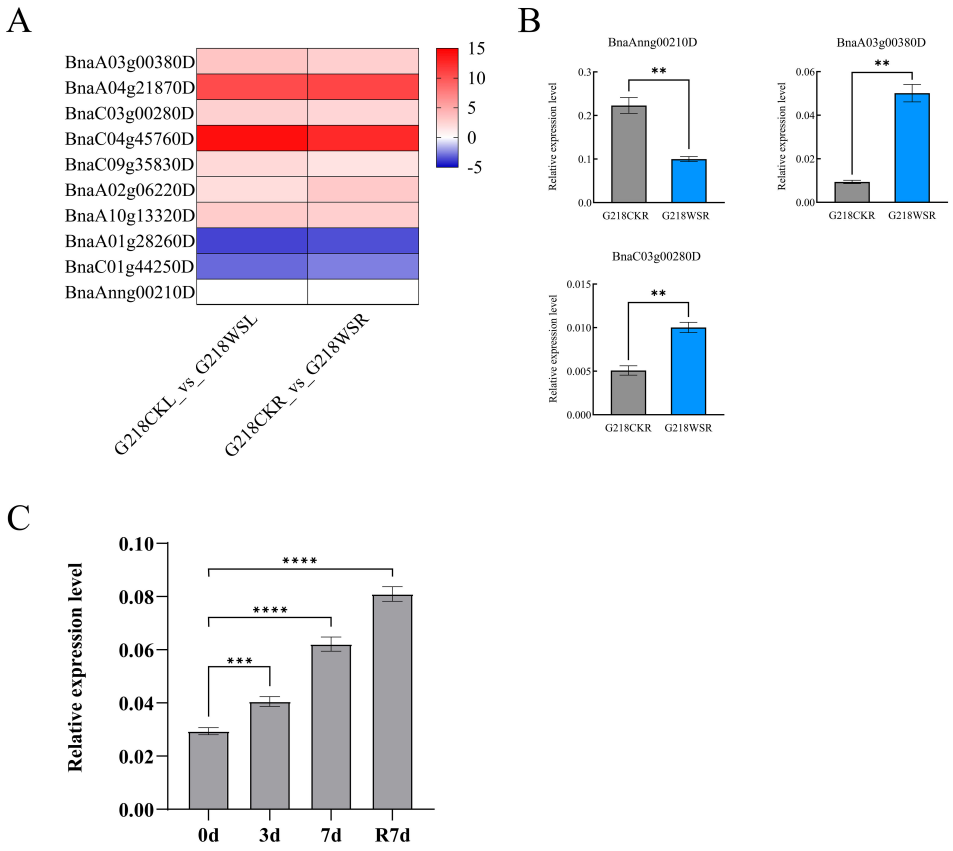


FIGURE 5
The *BnaPDX1.3* gene in *B. napus* is involved in the response to waterlogging stress. **(A)** RNA-seq analysis of *PDX* genes in *B. napus* G218 leaves and roots after 6 days of waterlogging stress. G218CKL denotes the control leaf, G218WSL the leaf after waterlogging stress, G218CKR the control root, and G218WSR the root after waterlogging stress. **(B)** qRT-PCR analysis of the *BnaPDX1.3* gene expression in *B. napus* G218 leaves after 6 days of waterlogging stress. The results of the qRT-PCR analysis are consistent with the RNA-seq. **(C)** qRT-PCR analysis of *BnaPDX1.3* gene expression in *B. napus* under waterlogging stress. qRT-PCR analysis of the *BnaPDX1.3* gene in *B. napus* at 0, 3, and 7 days of waterlogging stress, as well as after 7 days of reoxygenation following the end of stress. *p<0.05", "**p<0.01", ***p<0.001", ****p<0.0001".

expression of *BnaPDX1.3* in G218 seedlings significantly increased after waterlogging treatment and continued to rise even 7 days after reoxygenation (Figure 5C), indicating that *BnaPDX1.3* in *B. napus* responds positively to waterlogging stress.

4.3 *BnaPDX1.3* enhances waterlogging tolerance in *B. napus*

To investigate whether overexpression of the *BnaPDX1.3* gene in *B. napus* plays a positive regulatory role in response to waterlogging stress, two groups of *BnaPDX1.3*-overexpressing *B. napus* (PDX1.3#20 and PDX1.3#21) and WT *B. napus* at the three-leaf and one-heart stage were subjected to 15 days of waterlogging stress (Supplementary Figure S6). Under waterlogging stress, the wild-type leaves exhibited significant leaf abscission, turning purplish-red and yellowing. In contrast, the leaves of *BnaPDX1.3*-overexpressing rapeseed PDX1.3#20 showed less abscission, purplish-red coloration, and yellowing, while the leaves of *BnaPDX1.3*-overexpressing *B. napus* PDX1.3#21 did not exhibit obvious waterlogging phenotypes (Figure 6A). In terms of biomass, the overexpressing plants exhibited significantly higher values after waterlogging compared to the wild type (Figure 6B).

qRT-PCR analysis showed that the expression level of the *BnaPDX1.3* gene in PDX1.3#20 overexpressing plants was 19 times higher than that in the wild type, while in PDX1.3#21, it was 45 times higher. After 7 days of waterlogging stress, the expression levels of the *BnaPDX1.3* gene in both overexpressing and wild-type plants were significantly upregulated. PDX1.3#20 was upregulated by 3.4 times, and PDX1.3#21 by 3.2 times, while the wild type showed only a 2.2-fold upregulation. After 15 days of waterlogging stress, the expression levels of the *BnaPDX1.3* gene in all three groups were downregulated to levels lower than those before treatment. However, the transgenic plants still maintained relatively high expression levels (Figure 6C).

Under waterlogging stress, the vitamin B6 content in the overexpressing plants was significantly higher than that in the wild-type plants, peaking after 12 days of stress. The VB6 peak in PDX1.3#21 was 3.4 times higher than in the wild type, while the VB6 peak in PDX1.3#20 was 2.2 times higher (Figure 6D). In summary, plants with higher expression of the *BnaPDX1.3* gene synthesize more vitamin B6 and exhibit greater tolerance to waterlogging.

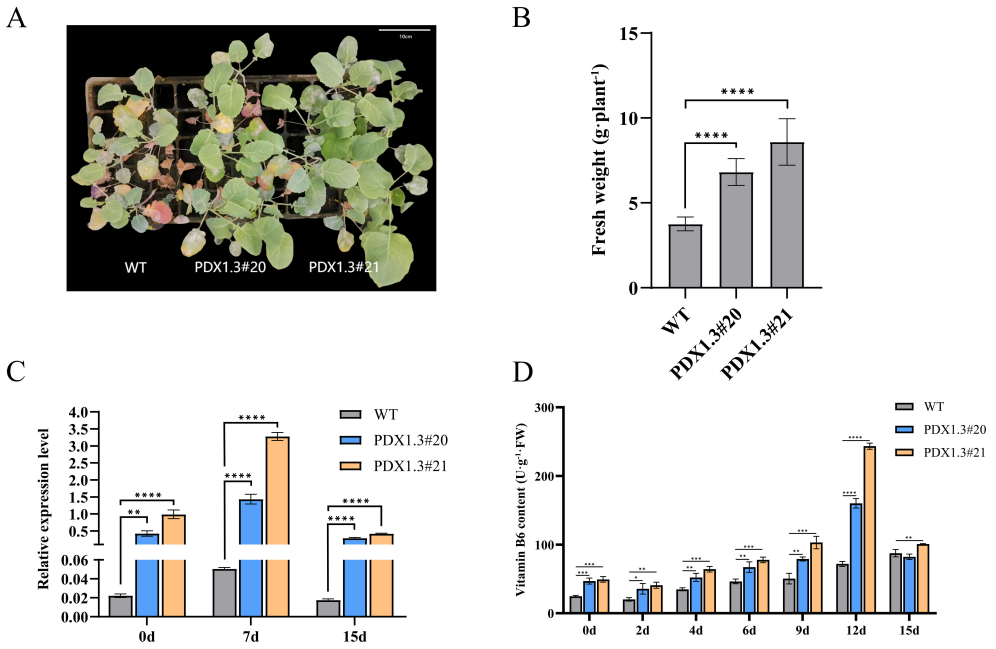


FIGURE 6
The *BnaPDX1.3* gene enhances waterlogging tolerance in *B. napus*. **(A)** Photographs of the *BnaPDX1.3* gene overexpressing *B. napus* and wild-type (WT) plants after 15 days of waterlogging stress. The photographs show that, compared to the overexpressing plants (PDX1.3#20 and PDX1.3#21), the wild-type plants exhibit more pronounced symptoms, including leaf shedding, purple-red discoloration, and yellowing. **(B)** Biomass analysis of the *BnaPDX1.3* gene overexpressing *B. napus* and wild-type plants after 15 days of waterlogging stress. The biomass of overexpressing plants was significantly higher than that of wild-type plants following waterlogging stress. **(C)** Quantitative expression analysis of the *BnaPDX1.3* gene in overexpressing and wild-type *B. napus* leaves under waterlogging stress. Overexpressing plants exhibited significantly higher expression levels of the *BnaPDX1.3* gene compared to wild-type plants. **(D)** Vitamin B6 (VB6) content in overexpressing and wild-type *B. napus* leaves under waterlogging stress. Plants with stronger *BnaPDX1.3* gene expression had higher VB6 content. This, along with their phenotype and biomass, corresponds to greater waterlogging tolerance. *p<0.05", "**p<0.01", ***p<0.001", ****p<0.0001".

4.4 Increased expression of *BnaPDX1.3* strengthens the plant antioxidant system

Upon detecting antioxidant enzyme activity in plant leaves under waterlogging stress, it was found that transgenic plants exhibited stronger antioxidant enzyme activity and more stable H_2O_2 content. During the early stages of waterlogging (2–4 days), the SOD activity in transgenic plants was significantly higher than that in wild-type plants. It gradually returned to levels similar to those in wild-type plants between 6 and 12 days, but remained significantly higher at 15 days, maintaining a high level of SOD activity (Figure 7A).

During the early stages of waterlogging (2–4 days), the CAT activity in transgenic plants was significantly higher than that in wild-type plants. It gradually decreased, but the trend was more stable in transgenic plants. At 15 days, CAT activity remained significantly higher in transgenic plants, maintaining a high level of CAT activity (Figure 7B).

From 2 to 6 days, the POD activity in transgenic plants was significantly higher than that in wild-type plants. During this period, all three groups of plants showed a gradual increase, reaching the same level at 9 days, after which the activity stabilized (Figure 7C).

After 4 days of waterlogging stress, the H_2O_2 content in transgenic plants rapidly decreased, while in wild-type plants, it rapidly increased, with their H_2O_2 content significantly higher than that of transgenic plants. At 6 days, the H_2O_2 content of wild-type plants reached its peak and remained elevated throughout the later stages of stress, while the H_2O_2 content in *BnaPDX1.3*-overexpressing plants had stabilized after 4 days and remained significantly lower than in wild-type plants, maintaining a low level throughout the later stages of stress (Figure 7D).

In summary, high-level expression of *BnaPDX1.3* helps plants establish a stronger and more efficient antioxidant system under waterlogging stress, effectively preventing the accumulation of ROS and subsequent oxidative damage.

5 Discussion

In China, *B. napus* is the second-largest oilseed crop, with a planting area of approximately 6.57 million hectares, accounting for 38.6% of the country's total oil crop production (Wang et al., 2022; Li and Wang, 2022). The Yangtze River Basin is the main producing area for *B.napus* in China, where 91% of the total oilseed rape in China is produced (Tao et al., 2015), *B. napus* production in the region primarily follows a rice–rapeseed rotation. The soil has a relatively high water-holding capacity, and the rapeseed growing period coincides with the rainy season in the basin, which is

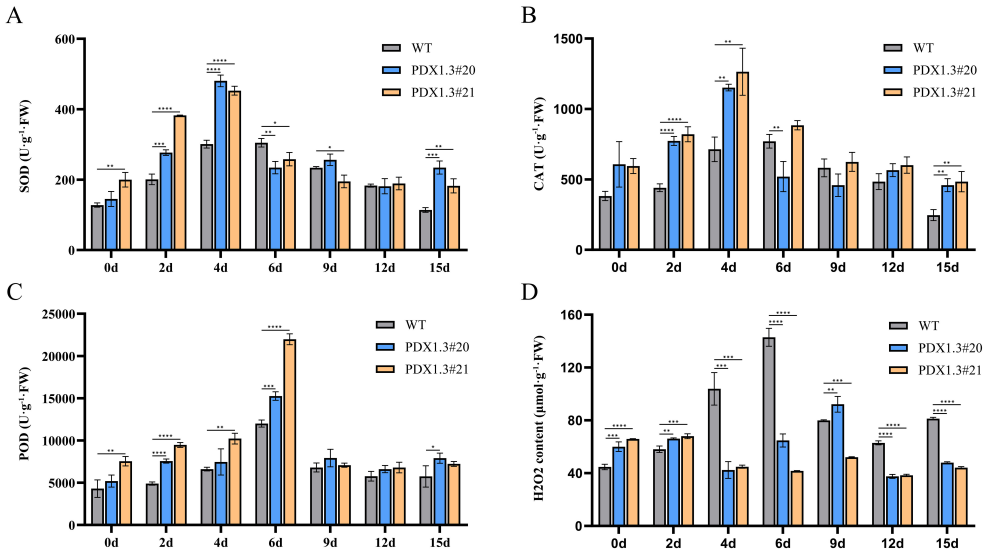


FIGURE 7
BnaPDX1.3 gene enhances antioxidant enzyme activity in *B. napus*. **(A)** SOD activity under waterlogging stress in *BnaPDX1.3*-overexpressing *B. napus* and wild-type plants. SOD activity in overexpressing plants shows a significant difference compared to wild-type plants under waterlogging stress. **(B)** CAT activity in *BnaPDX1.3*-overexpressing *B. napus* and wild-type plants under waterlogging stress. CAT activity in overexpressing plants is generally significantly higher than in wild-type plants under waterlogging stress. **(C)** POD activity in *BnaPDX1.3*-overexpressing *B. napus* and wild-type plants under waterlogging stress. Overexpressing plants exhibit stronger POD activity compared to wild-type plants under waterlogging stress. **(D)** H_2O_2 content in *BnaPDX1.3*-overexpressing *B. napus* and wild-type plants under waterlogging stress. H_2O_2 content in overexpressing plants is significantly lower and more stable compared to wild-type plants under waterlogging stress. *p<0.05", "**p<0.01", ***p<0.001", ****p<0.0001".

characterized by frequent high-intensity rainfall, making the region prone to waterlogging (Tao et al., 2024; Zhu et al., 2019). Waterlogging can affect root vitality and growth, photosynthetic efficiency, and the physiological metabolism of rapeseed, ultimately affecting yield and quality (Xu et al., 2018; Ploschuk et al., 2018).

Vitamin B6, as an endogenous growth regulator, is considered an antioxidant in plants, influencing physiological metabolism, growth, development, and stress resistance (Colinas et al., 2016; Moccand et al., 2014; Zhang et al., 2014; Huang et al., 2013). The *PDX* family genes, key players in the vitamin B6 synthesis pathway, have been shown to be involved in responding to various environmental stresses in plants. Studies have shown that the three *PDX1* and *PDX2* genes in *Arabidopsis thaliana* respond to strong light, cold, drought, and ozone stress (Bagri et al., 2018; Leasure et al., 2011). The *pdx1.3* mutant of *Arabidopsis thaliana* is more sensitive to salt, oxidative, and osmotic stress (Titiz et al., 2006; Moccand et al., 2014). Currently, research on plant *PDX* genes has primarily focused on *Arabidopsis thaliana*, and no studies have reported on the *PDX* genes in *B. napus*. As a gene involved in vitamin B6 synthesis, the effect of *BnaPDX1.3* on *B. napus's* response to waterlogging stress has not been fully elucidated. Therefore, this study aims to explore the role of the *BnaPDX1.3* gene in *B. napus's* response to waterlogging stress by analyzing the phenotype, as well as physiological and biochemical indicators, of T1 generation *BnaPDX1.3*-overexpressing rapeseed plants under waterlogging stress.

The research findings indicate that, after a 15-day waterlogging stress experiment, the transgenic rapeseed plants with higher expression levels of the *BnaPDX1.3* gene exhibited healthier leaves and greater biomass compared to wild-type plants, suggesting that the *BnaPDX1.3* gene enhances the waterlogging tolerance of *B. napus*. To delve deeper into the functional mechanism of the *BnaPDX1.3* gene in *B. napus*, this study examined the changes in VB6 content under waterlogging stress. The results revealed that the VB6 content in the transgenic plants was generally higher than that in the wild-type plants, with the peak VB6 content in the transgenic plants on day 12 being 3.4 times higher than that in the wild-type plants.

H_2O_2, as a plant stress signaling molecule and reactive oxygen species, can reflect the plant's ability to respond to and resist stress. In this study, the analysis of H_2O_2 content after waterlogging showed that the wild-type plants accumulated excessive H_2O_2 6 days after waterlogging and maintained high levels of H_2O_2 throughout the later stages of the stress, unable to effectively prevent cellular oxidative damage. However, the *BnaPDX1.3*-overexpressing plants maintained lower H_2O_2 levels under waterlogging conditions. The analysis of antioxidant enzyme activity indicated that the *BnaPDX1.3*-overexpressing plants had significantly higher SOD, POD, and CAT enzyme activities under waterlogging stress compared to the wild-type control plants, demonstrating a stronger antioxidant capacity. This effectively eliminates ROS generated by waterlogging, reduces oxidative damage, and thereby enhances plant tolerance to waterlogging.

We conclude that overexpressing plants, with high *BnaPDX1.3* gene expression, synthesize more VB6, exhibit stronger antioxidant enzyme activity, and possess a more efficient and stable ROS scavenging system, thereby demonstrating healthy growth under waterlogging stress. In summary, the *BnaPDX1.3* gene enhances the

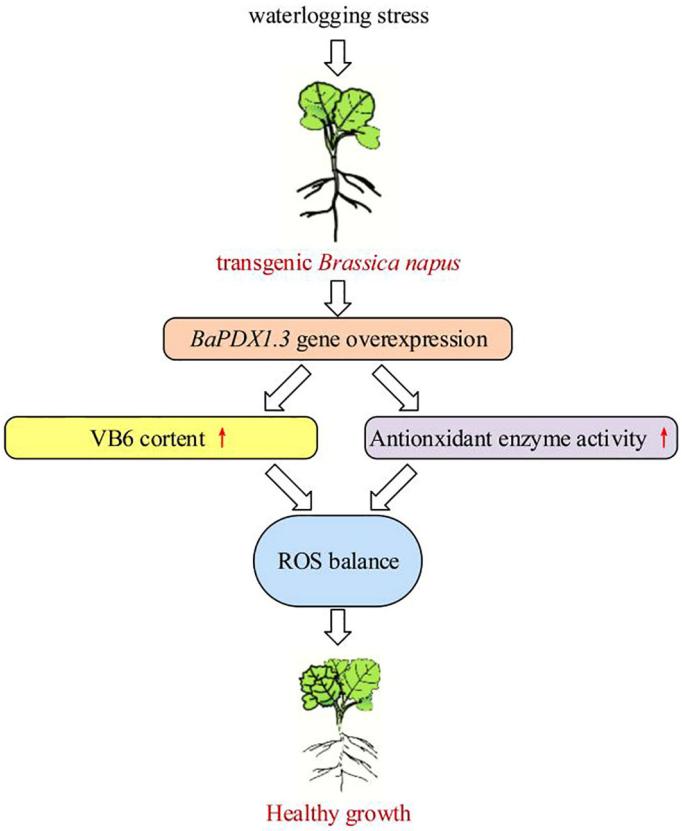


FIGURE 8
Impact of overexpression of *BnaPDX1.3* in *B. napus* on plant waterlogging tolerance. *BnaPDX1.3*-overexpressing *B. napus* plants exhibit high expression levels of the *BnaPDX1.3* gene under waterlogging stress, synthesize high levels of VB6, and significantly enhance antioxidant enzyme activity. This results in a stable dynamic balance of ROS under waterlogging stress, supporting healthy growth.

waterlogging tolerance of *B. napus*, which is of great significance for its response to waterlogging stress. The event model revealed in this study is shown in Figure 8. Additional studies have shown that VB6 can act as an antioxidant, participating in the scavenging of ROS accumulation (Lu et al., 2022; Danon et al., 2005; Zhang et al., 2022). Moreover, *pdx1* mutants impair IAA biosynthesis (Boycheva et al., 2015; Chen and Xiong, 2009), and impaired VB6 biosynthesis can inhibit ABA biosynthesis (Liu et al., 2022). In summary, VB6 participates in plant responses to adverse stress by regulating plant hormone synthesis and antioxidant systems, which may explain why overexpressing *BnaPDX1.3* plants exhibit stronger waterlogging tolerance.

6 Conclusion

The bioinformatics analysis of the *PDX* gene family and the waterlogging stress experiment analysis with the G218 wild type in this study indicate that the *BnaPDX1.3* gene is involved in the response of *B. napus* to waterlogging stress. To further investigate the regulatory role of the *BnaPDX1.3* gene in waterlogging stress, this study generated *BnaPDX1.3*-overexpressing *B. napus* plants through *Agrobacterium*-mediated transformation. After positive detection and screening, T1 generation positive seedlings were obtained. The T1 generation *BnaPDX1.3*-overexpressing *B. napus* plants were subjected to a 15-day waterlogging stress treatment. It was observed that *BnaPDX1.3*-overexpressing *B. napus* plants had healthier leaves and greater biomass compared to wild-type plants, exhibiting better waterlogging tolerance. Additionally, they displayed a stronger antioxidant system and a more stable ROS content under waterlogging stress. The results of this study indicate that the *BnaPDX1.3* gene can enhance the waterlogging tolerance of *B. napus*, which is of great significance for its response to waterlogging stress.

Data availability statement

The original contributions presented in the study are publicly available. This data can be found here: https://www.ncbi.nlm.nih.gov/, accession number PRJNA898876.

Author contributions

MY: Writing – original draft, Writing – review & editing. BH: Writing – original draft, Writing – review & editing. HJ: Writing – original draft, Writing – review & editing. CG: Writing – original draft, Writing – review & editing. MG: Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This research was funded by the Hunan Agriculture Research System (HARS-03).

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.2025.1533219/full#supplementary-material

References

Asensi-Fabado, M. A., and Munné-Bosch, S. (2010). Vitamins in plants: occurrence, biosynthesis and antioxidant function. *Trends Plant Sci.* 15, 582–592. doi: 10.1016/j.tplants.2010.07.003

Bagri, D. S., Upadhyaya, D. C., Kumar, A., and Upadhyaya, C. P. (2018). Overexpression of PDX-II gene in potato (Solanum tuberosum L.) leads to the enhanced accumulation of vitamin B6 in tuber tissues and tolerance to abiotic stresses. *Plant Sci.* 272, 267–275. doi: 10.1016/j.plantsci.2018.04.024

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

Boycheva, S., Dominguez, A., Rolcik, J., Boller, T., and Fitzpatrick, T. B. (2015). Consequences of a deficit in vitamin B6 biosynthesis *de novo* for hormone homeostasis and root development in Arabidopsis. *Plant Physiol.* 167, 102–117. doi: 10.1104/pp.114.247767

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, H., and Xiong, L. (2005). Pyridoxine is required for post-embryonic root development and tolerance to osmotic and oxidative stresses. *Plant J.* 44, 396–408. doi: 10.1111/j.1365-313X.2005.02538.x

Chen, H., and Xiong, L. (2009). The short-rooted vitamin B6-deficient mutant pdx1 has impaired local auxin biosynthesis. *Planta*. 229 (6), 1303–1310. doi: 10.1007/s00425-009-0912-8

Chung, K. R. (2012). Stress response and pathogenicity of the necrotrophic fungal pathogen alternaria alternata. *Scientifica* 2012, 635431. doi: 10.6064/2012/635431

Colinas, M., Eisenhut, M., Tohge, T., Pesquera, M., Fernie, A. R., Weber, A. P., et al. (2016). Balancing of B6 vitamers is essential for plant development and metabolism in arabidopsis. *Plant Cell* 28, 439–453. doi: 10.1105/tpc.15.01033

Danon, A., Miersch, O., Felix, G., Camp, R. G., and Apel, K. (2005). Concurrent activation of cell death-regulating signaling pathways by singlet oxygen in Arabidopsis thaliana. *Plant J.* 41, 68–80. doi: 10.1111/j.1365-313X.2004.02276.x

Dell'Aglio, E., Boycheva, S., and Fitzpatrick, T. B. (2017). The pseudoenzyme PDX1.2 sustains vitamin B6 biosynthesis as a function of heat stress. *Plant Physiol.* 174, 2098–2112. doi: 10.1104/pp.17.00531

Denslow, S. A., Rueschhoff, E. E., and Daub, M. E. (2007). Regulation of the Arabidopsis thaliana vitamin B6 biosynthesis genes by abiotic stress. *Plant Physiol. Bioch* 45, 152–161. doi: 10.1016/j.plaphy.2007.01.007

Dong, Y. X., Sueda, S., Nikawa, J., and Kondo, H. (2004). Characterization of the products of the genes SNO1 and SNZ1 involved in pyridoxine synthesis in Saccharomyces cerevisiae. *Eur. J. Biochem.* 271, 745–752. doi: 10.1111/j.1432-1033.2003.03973.x

Drewke, C., and Leistner, E. (2001). Biosynthesis of vitamin B6 and structurally related derivatives. *Vitam. Horm.* 61, 121–155. doi: 10.1016/s0083-6729(01)61004-5

Eliot, A. C., and Kirsch, J. F. (2004). Pyridoxal phosphate enzymes: mechanistic, structural, and evolutionary considerations. *Annu. Rev. Biochem.* 73, 383–415. doi: 10.1146/annurev.biochem.73.011303.074021

Fitzpatrick, T. B., Amrhein, N., Kappes, B., Macheroux, P., Tews, I., and Raschle, T. (2007). Two independent routes of *de novo* vitamin B6 biosynthesis: not that different after all. *Biochem. J.* 407, 1–13. doi: 10.1042/BJ20070765

Gorelova, V., Colinas, M., Dell'Aglio, E., Flis, P., Salt, D. E., and Fitzpatrick, T. B. (2022). Phosphorylated B6 vitamer deficiency in SALT OVERLY SENSITIVE 4 mutants compromises shoot and root development. *Plant Physiol.* 188, 220–240. doi: 10.1093/plphys/kiab475

Havaux, M., Ksas, B., Szewczyk, A., Rumeau, D., Franck, F., Caffarri, S., et al. (2009). Vitamin B6 deficient plants display increased sensitivity to high light and photo-oxidative stress. *BMC Plant Biol.* 9, 130. doi: 10.1186/1471-2229-9-130

Hong, B., Zhou, B., Peng, Z., Yao, M., Wu, J., Wu, X., et al. (2023). Tissue-specific transcriptome and metabolome analysis reveals the response mechanism of brassica napus to waterlogging stress. *Int. J. Mol. Sci.* 24, 6015. doi: 10.3390/ijms24076015

Huang, S., Zhang, J., Wang, L., and Huang, L. (2013). Effect of abiotic stress on the abundance of different vitamin B6 vitamers in tobacco plants. *Plant Physiol. Bioch* 66, 63–67. doi: 10.1016/j.plaphy.2013.02.010

Jeong, H., and Vacanti, N. M. (2020). Systemic vitamin intake impacting tissue proteomes. *Nutr Metab (Lond).* 17, 73. doi: 10.1186/s12986-020-00491-7

Leasure, C. D., Tong, H. Y., Hou, X. W., Shelton, A., Minton, M., Esquerra, R., et al. (2011). root uv-b sensitive mutants are suppressed by specific mutations in ASPARTATE AMINOTRANSFERASE2 and by exogenous vitamin B6. *Mol. Plant* 4, 759–770. doi: 10.1093/mp/ssr033

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

Leuendorf, J. E., Mooney, S. L., Chen, L., and Hellmann, H. A. (2014). Arabidopsis thaliana PDX1.2 is critical for embryo development and heat shock tolerance. *Planta* 240, 137–146. doi: 10.1007/s00425-014-2069-3

Li, Q., and Wang, C. (2022). An evaluation of chinese rapeseed production efficiency based on three-stage DEA and malmquist index. *Sustainability-Basel* 14, 15822. doi: 10.3390/su142315822

Liu, H., Lu, C., Li, Y., Wu, T., Zhang, B., Liu, B., et al. (2022). The bacterial effector AvrRxo1 inhibits vitamin B6 biosynthesis to promote infection in rice. *Plant Commun.* 3, 100324. doi: 10.1016/j.xplc.2022.100324

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, C., Tian, Y., Hou, X., Hou, X., Jia, Z., Li, M., et al. (2022). Multiple forms of vitamin B6 regulate salt tolerance by balancing ROS and abscisic acid levels in maize root. *Stress Biol.* 2, 39. doi: 10.1007/s44154-022-00061-2

Moccand, C., Boycheva, S., Surriabre, P., Tambasco-Studart, M., Raschke, M., Kaufmann, M., et al. (2014). The pseudoenzyme PDX1.2 boosts vitamin B6 biosynthesis under heat and oxidative stress in Arabidopsis. *J. Biol. Chem.* 289, 8203–8216. doi: 10.1074/jbc.M113.540526

Mooney, S., Leuendorf, J. E., Hendrickson, C., and Hellmann, H. (2009). Vitamin B6: a long known compound of surprising complexity. *Molecules* 14 (1), 329–351. doi: 10.3390/molecules14010329

Percudani, R., and Peracchi, A. (2003). A genomic overview of pyridoxal-phosphate-dependent enzymes. *EMBO Rep.* 4, 850–854. doi: 10.1038/sj.embor.embor914

Ploschuk, R. A., Miralles, D. J., Colmer, T. D., Ploschuk, E. L., and Striker, G. G. (2018). Waterlogging of winter crops at early and late stages: impacts on leaf physiology, growth and yield. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.01863

Raschle, T., Amrhein, N., and Fitzpatrick, T. B. (2005). On the two components of pyridoxal 5'-phosphate synthase from Bacillus subtilis. *J. Biol. Chem.* 280, 32291–32300. doi: 10.1074/jbc.M501356200

Ristilä, M., Strid, H., Eriksson, L. A., Strid, A., and Sävenstrand, H. (2011). The role of the pyridoxine (vitamin B6) biosynthesis enzyme PDX1 in ultraviolet-B radiation responses in plants. *Plant Physiol. Bioch* 49, 284–292. doi: 10.1016/j.plaphy.2011.01.003

Roje, S. (2007). Vitamin B biosynthesis in plants. *Phytochemistry* 68, 1904–1921. doi: 10.1016/j.phytochem.2007.03.038

Rueschhoff, E. E., Gillikin, J. W., Sederoff, H. W., and Daub, M. E. (2013). The SOS4 pyridoxal kinase is required for maintenance of vitamin B6-mediated processes in chloroplasts. *Plant Physiol. Bioch.* 63, 281–291. doi: 10.1016/j.plaphy.2012.12.003

Samsatly, J., Chamoun, R., Gluck-Thaler, E., and Jabaji, S. (2016). Genes of the *de novo* and Salvage Biosynthesis Pathways of Vitamin B6 are Regulated under Oxidative Stress in the Plant Pathogen Rhizoctonia solani. *Front. Microbiol.* 6. doi: 10.3389/fmicb.2015.01429

Subramanian, B., Gao, S., Lercher, M. J., Hu, S., and Chen, W. H. (2019). Evolview v3: a webserver for visualization, annotation, and management of phylogenetic trees. *Nucleic Acids Res.* 47, W270–W275. doi: 10.1093/nar/gkz357

Tambasco-Studart, M., Tews, I., Amrhein, N., and Fitzpatrick, T. B. (2007). Functional analysis of PDX2 from Arabidopsis, a glutaminase involved in vitamin B6 biosynthesis. *Plant Physiol.* 144, 915–925. doi: 10.1104/pp.107.096784

Tambasco-Studart, M., Titiz, O., Raschle, T., Forster, G., Amrhein, N., and Fitzpatrick, T. B. (2005). Vitamin B6 biosynthesis in higher plants. *Proc. Natl. Acad. Sci. U.S.A* 102, 13687–13692. doi: 10.1073/pnas.0506228102

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

Tao, R., Li, H., Lu, J., Bu, R., Li, X., Cong, R., et al. (2015). Crop rotation-dependent yield responses to fertilization in winter oilseed rape (Brassica napus L.). *Crop J.* 3 (5), 396–404. doi: 10.1016/j.cj.2015.04.007

Tao, Y., Li, D., Yu, Y., Lu, C., Huang, M., Wang, H., et al. (2024). Optimum transplanting date for rape forage and grain yields in the ridge culture place planting system on the yangtze river delta. *Appl. Sci.* 14, 3207. doi: 10.3390/app14083207

Titiz, O., Tambasco-Studart, M., Warzych, E., Apel, K., Amrhein, N., Laloi, C., et al. (2006). PDX1 is essential for vitamin B6 biosynthesis, development and stress tolerance in Arabidopsis. *Plant J.* 48, 933–946. doi: 10.1111/j.1365-313X.2006.02928.x

Vanderschuren, H., Boycheva, S., Li, K. T., Szydlowski, N., Gruissem, W., and Fitzpatrick, T. B. (2013). Strategies for vitamin B6 biofortification of plants: a dual role as a micronutrient and a stress protectant. *Front. Plant Sci.* 4. doi: 10.3389/fpls.2013.00143

Wagner, S., Bernhardt, A., Leuendorf, J. E., Drewke, C., Lytovchenko, A., Mujahed, N., et al. (2006). Analysis of the Arabidopsis rsr4-1/pdx1-3 mutant reveals the critical function of the PDX1 protein family in metabolism, development, and vitamin B6 biosynthesis. *Plant Cell* 18, 1722–1735. doi: 10.1105/tpc.105.036269

Wang, C., Xu, M., Wang, Y., Batchelor, W. D., Zhang, J., Kuai, J., et al. (2022). Long-term optimal management of rapeseed cultivation simulated with the CROPGRO-canola model. *Agronomy* 12, 1191. doi: 10.3390/agronomy12051191

Wetzel, D. K., Ehrenshaft, M., Denslow, S. A., and Daub, M. E. (2004). Functional complementation between the PDX1 vitamin B6 biosynthetic gene of Cercospora nicotianae and pdxJ of Escherichia coli. *FEBS Lett.* 564, 143–146. doi: 10.1016/S0014-5793(04)00329-1

Xu, J., Qiao, X., Tian, Z., Zhang, X., Zou, X., Cheng, Y., et al. (2018). Proteomic analysis of rapeseed root response to waterlogging stress. *Plants* 7, 71. doi: 10.3390/plants7030071

Zhang, Y., Jin, X., Ouyang, Z., Li, X., Liu, B., Huang, L., et al. (2015). Vitamin B6 contributes to disease resistance against Pseudomonas syringae pv. tomato DC3000 and Botrytis cinerea in Arabidopsis thaliana. *J. Plant Physiol.* 175, 21–25. doi: 10.1016/j.jplph.2014.06.023

Zhang, Y., Liu, B., Li, X., Ouyang, Z., Huang, L., Hong, Y., et al. (2014). The *de novo* biosynthesis of vitamin B6 is required for disease resistance against Botrytis cinerea in tomato. *Mol. Plant Microbe* 27, 688–699. doi: 10.1094/MPMI-01-14-0020-R

Zhang, C. Y., Peng, L. J., Chen, G. Y., Zhang, H., and Yang, F. Q. (2022). Investigation on the peroxidase-like activity of vitamin B6 and its applications in colorimetric detection of hydrogen peroxide and total antioxidant capacity evaluation. *Molecules* 27, 4262. doi: 10.3390/molecules27134262

Zhu, Y., Liao, S., Liu, Y., Li, X., Ren, T., Cong, R., et al. (2019). Differences of annual nutrient budgets between rapeseed– rice and wheat– rice rotations in the Yangtze River Basin. *J. Plant Nutr. Fertilizers* 25, 64–73. doi: 10.11674/zwyf.18031

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 06 March 2025
DOI 10.3389/fpls.2025.1545835

OPEN ACCESS

EDITED BY
Gurjeet Singh,
Texas A and M University, United States

REVIEWED BY
Vijay Rajamanickam,
Indian Agricultural Research Institute (ICAR), India
Archana Kumari,
YS Parmar University, Solan, India
Pooja Kanwar Shekhawat,
National Research Centre on Seed Spices (ICAR), India

*CORRESPONDENCE
Zahid Hussain Shah
zahid.uaar578@gmail.com

RECEIVED 23 December 2024
ACCEPTED 17 February 2025
PUBLISHED 06 March 2025

CITATION
Alghabari F and Shah ZH (2025) Deciphering salt tolerance mechanisms in synthetic hexaploid and bread wheat under humic acid application: physiological and genetic perspectives.
Front. Plant Sci. 16:1545835.
doi: 10.3389/fpls.2025.1545835

Deciphering salt tolerance mechanisms in synthetic hexaploid and bread wheat under humic acid application: physiological and genetic perspectives

Fahad Alghabari[1] and Zahid Hussain Shah[2]*

[1]Department of Arid Land Agriculture, King Abdulaziz University, Jeddah, Saudi Arabia, [2]Department of Plant Breeding and Genetics, PMAS-Arid Agriculture University, Rawalpindi, Pakistan

Salt stress is a potential constraint that perturbs plant physiological and osmolytic processes, and induces oxidative stress. The plant biostimulant, such as humic acid (HA) is capable to improve the wheat-tolerance to salt stress through triggering the plant defense mechanisms and regulating the genetic determinants. In this context the present study has comparatively evaluated the effect of HA on salt tolerant synthetic hexaploid (SH) and salt susceptible bread wheat (BW) genotypes. The experiment was performed in three replicates using randomized complete block design (RCBD) having two factorial arrangements, with HA treatment as one, while genotype as second factor. HA treatment significantly enhanced chlorophyll (33.33%–100%) and photosynthesis (31.25%–50%), and significantly reduced the glycine betaine (GB) (42.85%–77.77%), proline (20%–28.57%) and Na^+/K^+ ratio (33.33%–50%) in salt stressed SH and BW genotypes. Additionally, HA significantly increase the activities superoxide dismutase (SOD), peroxidase (POD), and catalase (CAT) by 57.14%–66.67%, 54.54%–83.33%, and 55.55%–80%, respectively in all salt stressed genotypes. The salinity associated genes *TaNHX1, TaHKT1,4, TaAKT1, TaPRX2A TaSOD* and *TaCAT1* were upregulated, while *TaP5CS* was downregulated in SH and BW genotypes corresponding to their regulatory traits. Furthermore, the multivariate analysis including correlation, principal component analysis (PCA) and heatmap dendrogram further rectified the strong impact of HA on the strength of association and expression of stress marker traits. Overall, the SH genotypes showed more strong response to the HA and illustrated significant tolerance to salt stress based upon physiological, biochemical and genetic indicators. Conclusively, the SH can serve as a bridge to transfer alien genes associated with salt tolerance into elite bread wheat germplasm.

KEYWORDS

synthetic hexaploid, gene regulation, antioxidant, correlation, heatmap

1 Introduction

Wheat is an important cereal crop dominating the most of the arable land worldwide. The stressful environment can decrease wheat production by 6% as predicted by different climatic models (Teleubay et al., 2024). Salt stress imposes negative impacts on 20% of world arable land owe to anthropogenic activities and the change in climate (El Sabagh et al., 2021). Approximately 1.4 billion hectare area of the world is effected by salinity (FAO, 2024). The saline environment imposes osmotic stress, ionic imbalance and toxicity, through high Na^+ influx that increases Na^+/K^+ ratio and retards the plant vital processes and crucial functions (Shah et al., 2017). Besides, salt stress brings variation in ultrastructural components of cells, damages the membrane, perturbs the photosynthesis machinery, enhance the oxidative stress through reactive oxygen species (ROS) which limits the plant ultimate productivity (Sachdev et al., 2021). In depth understanding of stress tolerance in plants with respect to physiological mechanisms is an important breeding goal (Chaudhary et al., 2024). Therefore, the concept of physiological, biochemical and genetic mechanisms of wheat responses under salt stress is inevitable. Among physiological features, the loss of photosynthetic pigments, while among biochemical features the accumulation of osmolytes is the focal point of plant research particularly under salt stress (Gupta and Huang, 2014). Moreover, plants exposed to salt stress undergo various oxidative changes like production of reactive oxygen species (ROS), lipid-peroxidation and plasma membrane deterioration (Hasanuzzaman et al., 2021). Being living body plant try to retain its homeostasis through activating the antioxidant system having enzymes such as peroxidases (PODs), superoxide dismutases (SODs) and catalases (CATs) (Mehla et al., 2017). Therefore, an understanding of physiological and biochemical patterns of salinity tolerance can be used as a selection parameter for improving the wheat adaptability in saline environment (Irshad et al., 2022). In past many studies have proved strong link between physio-chemical and genetic responses of wheat under different kinds of abiotic stresses (Yang et al., 2014; El Sabagh et al., 2021; Alsamadany et al., 2023; Alghabari and Shah, 2024). Briefly, plant behave like a composite system to sustain their physio-chemical equilibrium through modulating their osmotic and genetic responses when exposed to salt stress (Shah et al., 2017). Besides, for comprehensive elucidation of salt tolerance mechanisms in wheat, it is important to explore the physio-chemical responses of plant in association with their genetic determinants (Gupta and Huang, 2014). For instance, the upregulation of genes *TaCAT1, TaSOD* and *TaPRX2A* in salt stressed wheat plants triggers the antioxidant activities and increases the ROS scavenging via activating the antioxidant defense system comprising the enzymes CAT, SOD and POD (Alghabari and Shah, 2024). Similarly, *TaP5CS* gene increases the proline accumulation in wheat during salt, drought and heat stresses (Aycan et al., 2022). High affinity K^+-transporters (*HKTs*), and Na^+ proton-exchangers (*NHXs*) are membrane transporters that respectively regulate K^+-influx and Na^+-efflux, therefore reducing the Na^+ toxicity through decreasing Na^+/K^+ ratio (Hussain et al., 2021). Likewise, an inward-rectifying K^+-channel, *AKT1* increases plant tolerance against osmotic-stress by increasing the concentration of K^+ in cell as reported by Ahmad et al. (2016). Salinity stress impairs wheat growth and yield by causing osmotic stress, ionic imbalance, and osmotic damage (Ali et al., 2022). HA overcomes these effects by increasing chlorophyl, photosynthesis and antioxidant enzyme activities as reported by Tahir et al. (2022) and Meng et al. (2023). Furthermore, it also regulates the osmolytic level, improves the osmotic adjustments and salt tolerance (Zhang et al., 2024). Besides, HA improves the Na^+/K^+ ratio in salt stressed plants. Meng et al. (2023) reported that HA reduces the proline accumulation in salt stressed plants, indicating that HA alleviates the osmotic stress, hence decreasing the need of proline as an osmoprotectant. Likewise, HA application is associated with decrease in GB content salt stressed plants, suggesting improvement in water status and homeostasis, thereby minimizing the necessity of GB accumulation (Carillo et al., 2008). Various studies in past have been conducted to elucidate the potential role of HA in enhancing the salt tolerance of wheat (Carillo et al., 2008; Ali et al., 2022; Tahir et al., 2022; Meng et al., 2023). The plant biostimulants such as humic acid (HA) plays executive role in maintaining plant physiological activities during salt stress through increasing the production of osmoprotectants, and through regulating the expression of genes imparting stress tolerance (Ahmad et al., 2022; Meng et al., 2023). Reduced sodium toxicity, osmoprotectant regulation, a decrease in the Na+/K+ ratio, and an increase in antioxidant enzyme activity are all linked to HA-induced salt tolerance, which in turn lowers oxidative stress levels (Shukry et al., 2023). Although various studies have explored the aspects of salinity tolerance by integrating the physiological, biochemical, and genetic mechanisms, the present study aimed to elucidate the impact of HA on synthetic hexaploid (SH) wheat as compared to bread wheat (BW). Furthermore, this study supersedes the past researches by not only exploring the physiological, biochemical and genetic indices of salt tolerance but also providing a comparative analysis between wheat types, thereby providing new perceptions into their varying responses to HA under salt stress.

2 Materials and methods

The present experiment was performed within the glass house situated at the research site of the King Abdulaziz University (KAU), Jeddah, Saudi Arabia. The wheat genotypes shown in Table 1, including salt susceptible bread wheat cultivars, Kohistan-97, Fareed-06, A.Sattar (Irshad et al., 2022) and salt tolerant synthetic hexaploid (Li et al., 2018) were collected from Pakistan and evaluated for physiological, biochemical and genetic performance under saline condition in the presence and absence of HA. The experiment was executed in two factorial tri-replicated RCBD with HA and control treatments as one, and genotypes as another factor.

2.1 Plant growth and treatment application

The tri-replicated experiment was optimized in glass house (25/15 ± 2°C day/night temperature, 10 hour light-period, 60%

TABLE 1 Wheat germplasm containing synthetic hexaploid (SH) and salt susceptible bread wheat (BW) genotypes.

Genotype	Pedigree
Synthetic Hexaploids (Salt tolerant)	
SH1	TC870344/GU1//TEMPORALERA M 87/AGR/3/WBLL1
SH2	GPO8 KAZAKSTAN 6 WM98-99/4/KAUZ//ALTAR 84/AOS/3/KAUZ/5//KAUZ//ALTAR 84/AOS/3/KAUZ
SH3	CROC_1/AE.SQUAROSSA (205)//BORL95/3/KENNEDY
SH4	PAM94/3/ALTAR84/AEGILOPS SQUAROSA(TAUS)//OPATA/PASTOR
Bread Wheat (Salt suceptible)	
Kohistan-97	V-1562//CHIROCA(SIB)/(SIB)HORK/3/KUFRA-I/4/CARPINTERO(SIB)/(SIB)BLUEJAY[1964][2857]
Fareed-06	PTS/3/TOB/LFN//BB/4/BB/HD-832-5//ON
A. Sattar	PRL/PASTOR//2236(V.6550/Sutlej-86

humidity) using 2.5liter pots, having 30cm depth and 25 cm diameter. The pots were filled with natural soil (loamy soil with pH_6.5, EC 1.5 dS/m) and sown with six seeds, while each pot was left with four plants at seedling stage. The experimental set of pots was provided with HA (Dalian Vic Co., Ltd., Dalian, China) having specified composition (C=55%, P=0.01%, N=4.87%, K=11.2%, Ca=0.50%, Mg=0.22%). The HA was applied at 2 g kg^{-1} before sowing of seeds (Khaled and Fawy, 2011; Canellas et al., 2015). The pots were watered twice a week by 500 ml of 25% Hoagland nutrient solution (Sigma-Aldrich, USA). The pot was watered with150 mM of NaCl solution that set the EC value at 8.5 dS/m. The EC values were monitored on daily bases using the conductivity meter.Both control and experimental set of plants were subjected to salt stress at tillering stage 5 (Feekes growth scale) until the symptoms of saline stress indicators for instance leaf wilting, rolling and chlorosis. Besides, each treatment comprised of five pots, while for data analysis five plants were selected on random basis for each treatment of every replicate.

2.2 Quantification of physiological and osmolytic traits

The SPAD-502plus apparatus was used to quantify the chlorophyll content, while the photosynthesis rate was recorded using infra-red gas analyzer (IRGA) apparatus (ADC Bioscientific, UK) from attached flag leaf between 9am to 10am. The Na^+ and K^+ content were estimated according to the procedure described by Havre (1961). In short, the leaf samples were oven dried and converted to fine powder. Afterward, 0.5g of each sample was homogenized in 3ml $HClO_4$ and 8ml Conc HNO_3 and retained for 12h. The mixture was burnt at 300°C for 3h and double distilled water was supplemented till 50ml volume. The flame photometer was used to record the concentrations of Na^+ and K^+. Before taking measurements, the flame-photometer was standardized and properly calculated by using known K^+ and Na^+ standard solutions prepared from high purity KCl and NaCl. With a blank solution the instrument was set at zero, and a calibration curve was structured by recording the emission intensities of standard solutions. After calibration the flame photometer was used to measure the Na^+ and K^+ concentrations from prepared samples, and Na^+/K^+ ratio was estimated accordingly. Furthermore, the UV-Vis spectrophotometer (N5000, Shanghai, China) was used to record the proline concentration based on its reactivity with ninhydrin as explained by Carillo and Gibon (2011). The glycine betaine (GB) content was determined using the HPLC apparatus (Shimadzu Corp, Japan) following the protocol explained by Grieve and Grattan (1983).

2.3 Quantification of antioxidant activity

For the quantification of antioxidant enzyme activity, 2g of frozen leaf sample was homogenized in 2mL of 0.1M Tris-HCl (pH=7.4) at ice cold stage. The mixture was optimized at 4°C and subjected to centrifugation for 20 minutes at 20000 g. The supernatant attained was used to record the enzymatic activity based on the standard protocol explained by Alici and Arabaci (2016). The activity of superoxide dismutases (SODs) was assessed against the standard absorption curve of 560nm based upon its tendency to inhibit the photo-reduction of nitroblue-tetrazolium. Moreover, the activity of peroxidases (PODs)was recorded by spectrophotometer against the standard absorption curve of 420 nm using 4-methylcatechol as substrate. The catalases (CATs) activity was measured using spectrophotometer at 25°C against standard absorption curve of 240nm, using H_2O_2 as substrate. Moreover, the catalytic activities of all antioxidant enzymes were measured in enzyme units ($U\ mg^{-1}$*of protein).*

2.4 Gene expression analysis

The gene expression analysis was performed when plants showed the salt stress symptoms. The leaf samples from selected plants were instantaneously froze in a container having liquid nitrogen and stored at -80°C until the start of extraction process. The RNeasy kit (Qiagen, Germany) was used for RNA extraction according to the instructions of manufacturer. Besides, cDNA library was constructed from 2 µg RNA using QuantiTect reverse transcription kit (Qiagen, Germany). Furthermore, the transcript levels of salinity related genesin leaf tissue were quantified in qRT-PCR using SYBR Green Kit (BioFACT, Korea), while *TaActin1*-expressing gene was used for the expression normalization. Each expression profile from qRT-PCR analysis was analysed and confirmed following the procedure opted by Ahmed et al. (2022) using each of three biological and technical replicates. Furthermore, double-delta Ct values were used for the calculation of relative gene expression from each sample. The list of genes associated with salinity tolerance in wheat along with their forward and reverse primers is indicated in Table 2.

TABLE 2 The list of primers used in relative expression analysis of salt stress related genes in synthetic hexaploid (SH) and bread wheat (BW) genotypes.

Gene	Primer sequence	Reference
TaAKT1	CGGATAATGCCGTGAATG (F) TTATACTATCCTCCATGCCT (R)	Zeeshan et al., 2020
TaCAT1	TCTCTCGGCCAGAAGCTCG (F) AGGGAAGAACTTGGACGGC (R)	Al-Ashkar et al., 2019
TaPRX2A	CGTGTGTGTGATCATCAGTAAC (F) AGGCGGAGTGTAAATTACAAGA (R)	Su et al., 2020
TaHKT1,4	AGCAAGCTGAAGTTGAGGGG (F) AGAGTTGTGACAGAGCCGTG (R)	Dave et al., 2022
TaP5CS	GAAGGCTCTTATGGGTGTACTCAA (F) TAAAAGACCTTCAACACCCACAGG (R)	Aycan et al., 2022
TaSOD	TCCTTTGACTGGCCCTAATG (F) CTTCCACCAGCATTTCCAGT (R)	Jiang et al., 2019
TaNHX1	TGACGGAGGCAGAAGACCG (F) CCCAAAACTCTACACAGCGT (R)	Al-Ashkar et al., 2019

2.5 Data analysis

The Analysis of variance (ANOVA) was carried out using a computer-based statistics tool, Statistix ver. 8.1 (Mcgraw-Hill, 2008), at the probability level of 5% and LSD was computed, while the RStudio version 1.1.456 (RStudio Team, 2020) was applied to compute correlation, to perform principal component analysis (PCA), and to construct heatmap clusters based upon similarity matrix. The PCA biplots were constructed using "factoextra" and "FactoMineR", while the "GGally" and "ggplot2" were used to form Pearson's correlation chart. The packages "pheatmap" and "complex Heatmap" were used to form dendrogram.

3 Results

3.1 Chlorophyll, photosynthesis and osmolytes

The application of HA has significantly ($p \leq 0.05$) improved the chlorophyl and photosynthesis rate (Pn) in all salt stressed wheat genotypes as compared to no application of HA (Figure 1). The synthetic hexaploid (SH1 to SH4) depicted superior physiological performance under HA application compared to other bread wheat (BW) cultivars (Kohistan-97, Fareed-06, A. Sattar). The chlorophyll content illustrated significantly higher baseline in SH1 to SH4 and increased 21.4%-50%, whereas Kohistan-97 showed a highly variable response with an 80% rise in chl content. Besides, photosynthesis rate (Pn) increased by 28%-36% in SHs (SH1 to SH4) that revealed more stable and higher assimilation of CO_2 compared to sharper and inconsistent increase (42.8%-50%) in BW cultivars. Besides, under HA application, all salt stressed wheat genotypes showed significant ($p \leq 0.05$) decline in their proline and glycine betaine (GB) contents as compared to set of wheat genotypes grown under control condition (Figure 1). Conversely the proline accumulation declined from 28.5%-50%, while the GB content declined from 50%-72%, in salt stressed SHs (SH1 to SH4) due to HA as compared to control. Moreover, unlike Kohistan-97, Fareed-06, and A. Sattar, which showed dramatic fluctuations, the SH1 to SH4 kept higher and stable physiological responses and less variation in osmoprotectants (proline and GB) under salt stress due to application of HA.

3.2 Antioxidant enzymes and Na^+/K^+ ratio

The HA caused significant ($p \leq 0.05$) reduction in Na^+/K^+ ratio in all salt exposed set of plants as compared to the control set of plants (Figure 2). The SHs (SH1 to SH4) depicted significant reduction in Na^+/K^+ ratio 30% to 33.3% than BW cultivars Kohistan-97, Fareed-06 and A.Sattar 15% to 18.1% under saline conditions due to the application of HA, suggesting better ionic regulation and salt stress mitigation. Additionally, HA stimulated the activity of antioxidant enzymes catalase (CAT), superoxide dismutase (SOD), and peroxidase (POD) in all salt stressed wheat genotypes as compared to zero treatment of HA (Figure 2). The antioxidant enzymes superoxide dismutase (SOD) demonstrated increased catalytic activity by 20.8%-25%, in SH1 to SH4 as compared to 20% 26.6% in BW cultivars, suggesting their stable antioxidant defense system. Besides, the peroxidase (POD) activity increased by 27.2%-50% in SHs (SH1 to SH4) indicating better ROS scavenging compared to 42.8%-60% in BW genotypes. Furthermore, the catalase activity increased by 25%-44.4% in SH1 to SH4 explicating the better oxidative control under salt stress due to HA. Compared to Kohistan-97, Fareed-06, and A. Sattar, which manifested high variation in enzymatic activities, the SHs (SH1 to SH4) kept better Na^+/K^+ balance, consistent antioxidant response, and superior adaptability against salt stress under HA application.

3.3 Multivariate analysis

The correlogram revealed significant correlation among chlorophyll, Pn, proline, Na^+/K^+, GB, CAT, POD and SOD in all salt stressed wheat genotypes (Figure 3). The proline, GB and Na^+/K^+ illustrated showed significant positive paired association with one another, and showed significant negative paired association with antioxidant enzymes (CAT, POD and SOD), chl and Pn (Figure 3). Conversely, the antioxidant enzymes (CAT, POD and SOD), chl and Pn varied proportionally and illustrated paired association in same direction (Figure 3). Generally, the correlation analysis has proved strong paired association among all traits, both under control and HA treatments (Figure 3). Besides, correlogram further confirm the extent of association of traits significantly changed owe to HA application as compared to zero HA application under salt stress.

Besides, principal component analysis (PCA) further endorsed the results from correlation analysis. The biplot of both PCAs under HA and control treatment indicated the different orientation of trait

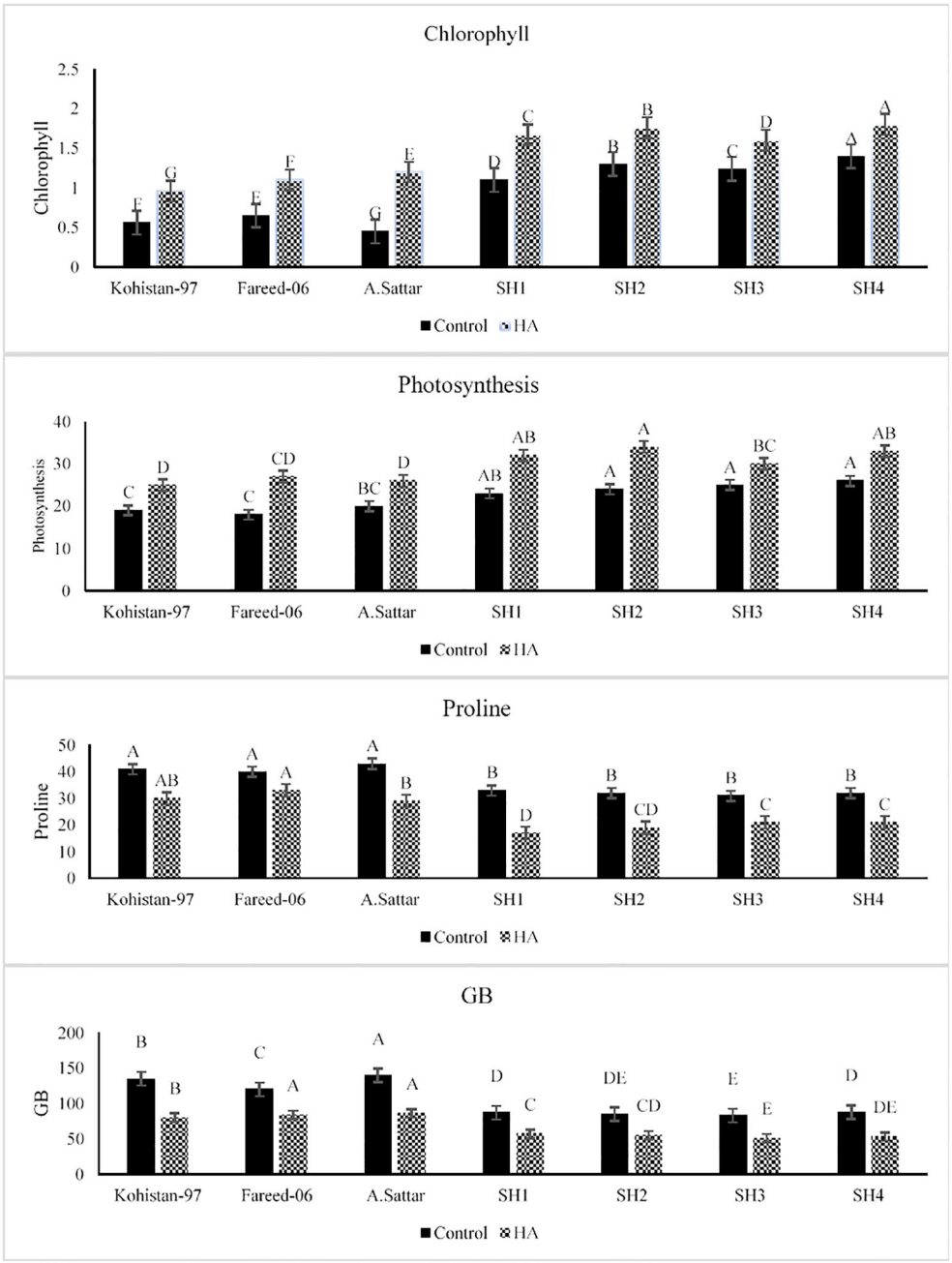


FIGURE 1
Effect of HA and control treatment on chlorophyll, photosynthesis, proline and glycine betaine content of different wheat genotypes grown under salt stressed environment. Chl, chlorophyll; Pn, photosynthesis; GB, glycine betaine. Graph bars represent the mean values of traits, analysed during tri-replicated two factorial experiment at $p \leq 0.05$. The bar values following the different letters are significantly different at $p \leq 0.05$.

vectors, which confirmed that association and expression of chl, Pn, proline, GB, Na^+/K^+ and antioxidant enzymes changed significantly due to HA as compared to control treatment under saline environment (Figure 4). Furthermore, PCA confirmed the differential response of each genotype in terms of trait association and expression due to HA application as compared to control as illustrated by their dispersed distribution within PCA biplot (Figure 4). In biplot, the SH (SH1 to SH4) genotypes were clustered at one quadrant, while salt susceptible BW genotypes were clustered on another quadrant of biplot. This indicated clear difference in the response of both groups of genotypes to HA under saline environment in terms of trait association (Figure 4). Furthermore, the varying impact of HA and control treatment on trait responses were further confirmed by differential positioning of ellipses on PCA biplot (Figure 5). The traits including Chl, Pn and antioxidant activity (SOD, POD and CAT) were more strongly expressed under HA treatment, while proline, GB and Na^+/K^+ were more strongly expressed under control treatment (Figure 5). The PCA biplot indicated that Dim 1 (93.8%) captured most of the variation, isolating the HA and control treatments, while Dim2

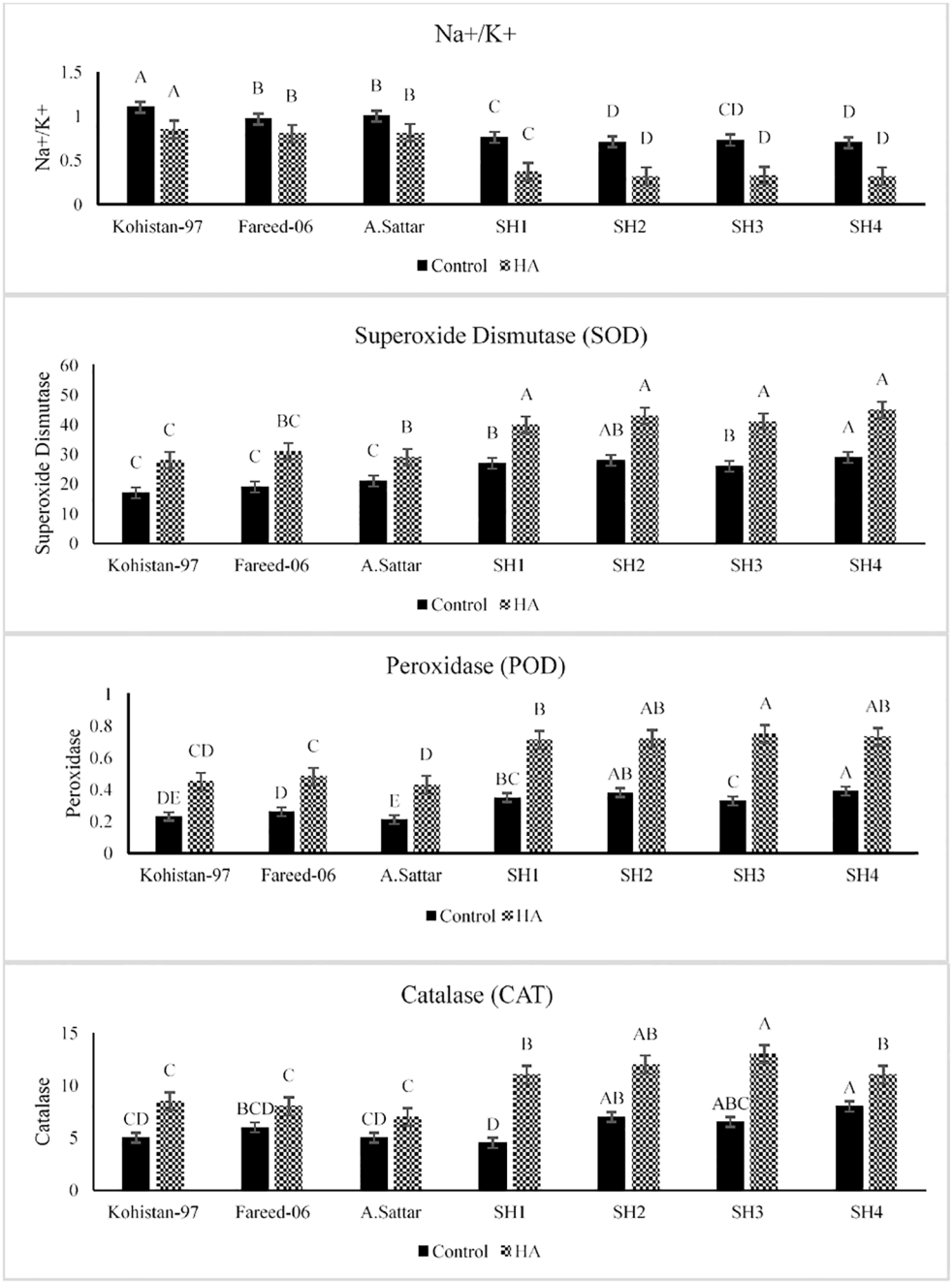


FIGURE 2
Effect of HA and control treatment on the activities of antioxidant enzymes (SOD, POD, and CAT) and Na^+/K^+ ratio of different wheat genotypes grown under salt stressed environment. SOD, superoxide dismutase; POD, peroxidase; CAT, catalase. Graph bars represent the mean values of traits, analysed during tri-replicated two factorial experiment at $p \leq 0.05$. The bar values following the different letters are significantly different at $p \leq 0.05$.

(2.9%) contributed slightly, suggesting collectively 96.7% of total variation. Although ellipses showed some overlap, but overall, the treatments showed distinct biochemical responses. Similarly, the different positioning of the ellipses representing the SH (SH1 to SH4) and susceptible BW (Kohistan-97, Fareed-06 and A. Sattar) in PCA biplot has further differentiated the responses of both group of genotypes with respect to the expression of traits (Figure 6). The PCA biplot indicated that that Dim1 (93%) captured most of the variation, while Dim2 (3%) added a small part of variation, together comprising the 96% of the total variation. Besides, the merged ellipses of SH group indicated the analogous response of SH (SH1 to SH4) genotypes while the merged ellipses of susceptible BW indicated the analogous response of BW (Kohistan-97, Fareed-06 and A. Sattar) cultivars in terms of trait expression and association. The traits including Chl, Pn and antioxidant activity (SOD, POD andCAT) were more strongly expressed in salt tolerant SH (SH1 to SH4), while proline, GB and Na^+/K^+ were more strongly expressed in susceptible BW (Kohistan-97, Fareed-06 and A. Sattar)

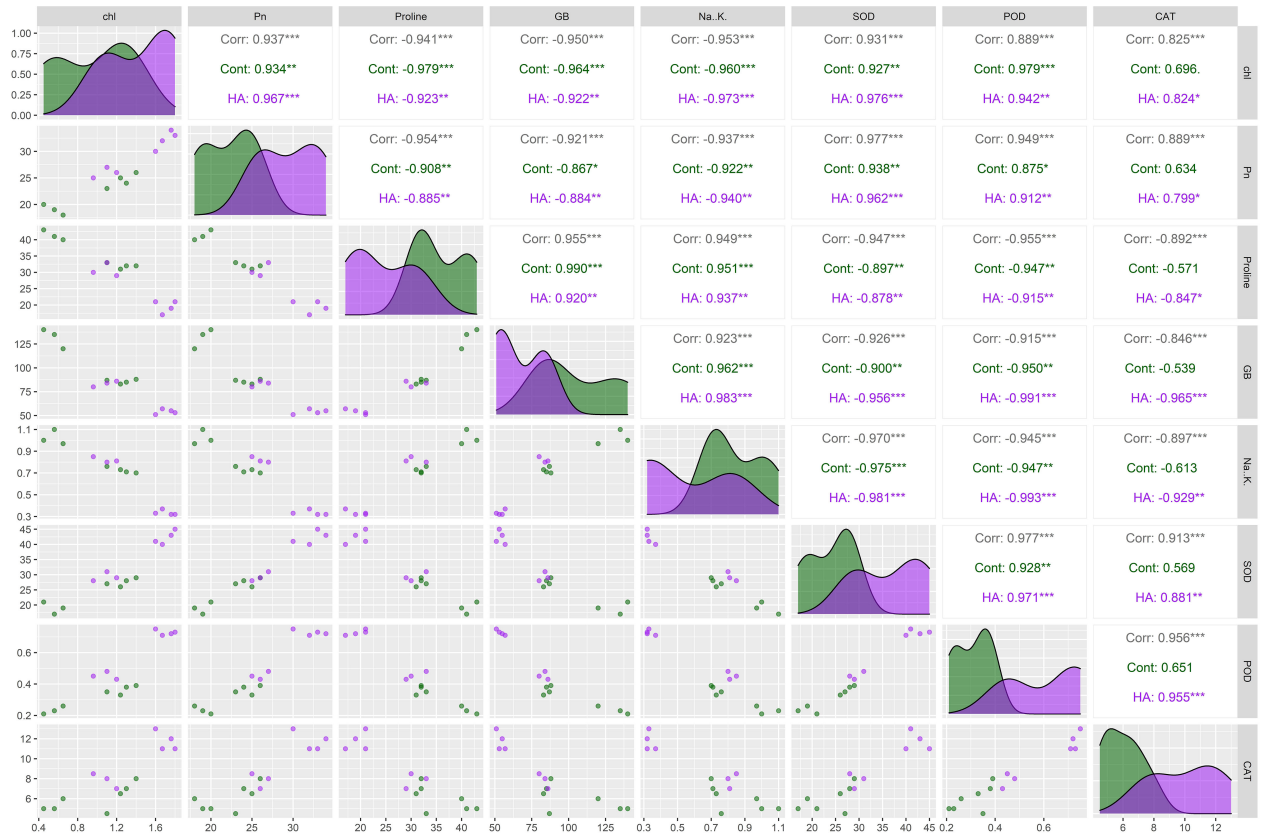


FIGURE 3
Correlogram showing the effect of control and HA treatment on the paired association of chlorophyll, photosynthesis, osmolytes (Proline and GB), Na^+/K^+ and antioxidant enzymes (SOD, POD and CAT) in salt stressed wheat genotypes. Chl, chlorophyll; Pn, photosynthesis; GB, glycine betaine; SOD, superoxide dismutase; POD, peroxidase; CAT, catalase. ***Significant at $p \leq 0.001$, **Significant at $p \leq 0.01$.

genotypes (Figure 6). The heatmap clustered dendrogram further endorsed the findings from PCA (Figure 7). The SH and BW genotypes showed varying trait response under control and HA application as indicated by varying banding pattern. Corresponding to the PCA results the heatmap has divided SH and susceptible BW genotypes into two different groups.

3.4 Gene expression analysis

The gene *TaP5CS* significantly ($p \leq 0.05$) downregulated in all salt stressed wheat genotypes due to HA (Figure 8). Among all genotypes, the SH (SH1 to SH4) wheat lines showed comparatively low transcript level (1.3 to 2 fold) of TaP5CS as compared to susceptible BW (Kohistan-97, Fareed-06 and A. Sattar) genotypes (Figure 8). Besides, the downregulation of *TaP5CS* was in accordance to the declined amount of proline in wheat genotypes grown in saline environment under the application of HA.

The genes *TaNHX1*, *TaHKT1,4* and *TaAKT1* regulating the K^+-influx and Na^+- compartmentalization were overexpressed in all salt stressed wheat genotypes from 2 to 4.2 fold, 2.1 to 4 fold, and 1.75 to 3.1 fold respectively, due to the application of HA (Figure 8). Overall, the SH (SH1 to SH4) wheat genotypes recorded significantly ($p \leq 0.05$) high expression of these genes, that imparted high influx of K^+, and high efflux of Na^+, as illustrated by their low Na^+/K^+ ratio as compared to susceptible BW (Kohistan-97, Fareed-06 and A. Sattar) genotypes under the application of HA.

The HA significantly ($p \leq 0.05$) triggered the expression of *TaSOD* (1.75 to 2.2 fold), *TaCAT1* (1.3 to 2.5 fold) and *TaPRX2A* (1.2 to 1.75 fold) in all salt stressed wheat genotypes (Figure 8). Overall, the relative expression of these genes was significantly high in SH (SH1 to SH4) as compared to susceptible BW genotypes. Besides, the increase in transcript level of *TaSOD, TaCAT1* and *TaPRX2A* was in agreement with increasing activity of antioxidant enzymes SOD, CAT and POD.

4 Discussion

Salinity is a potential constraint restricting wheat production throughout the world. Salt stress effect plant directly by perturbing physiological and biochemical processes and indirectly by imposing oxidative stress. However, plant biostimulant such as HA induces salt tolerance in plants owe to its potential role in decreasing Na^+ toxicity, enhancing osmolytic concentrations, lowering Na^+/K^+ ratio, and increasing ROS scavenging by speeding up the catalytic activities of antioxidant enzymes. Salt stress significantly decreases

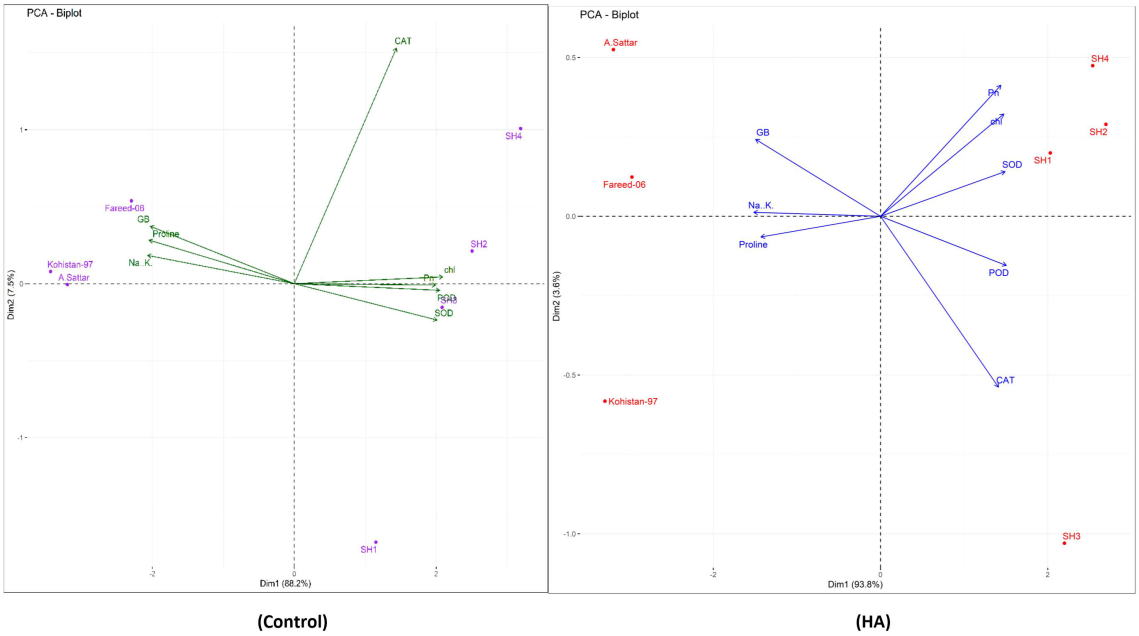


FIGURE 4
The PCA biplot based upon correlation matrix, indicating the effect of humic HA and control treatment on chlorophyll, photosynthesis, osmolytes (Proline and GB), Na^+/K^+ and antioxidant enzymes (SOD, POD and CAT) in salt stressed wheat genotypes. The orientations and closeness of traits vectors differs under HA application as compared to control treatment which confirms the changing association and expression of traits with respect to genotypes categorized into different biplot quadrants. Chl, chlorophyll; Pn, photosynthesis; GB, glycine betaine; SOD, superoxide dismutase; POD, peroxidase; CAT, catalase.

the chlorophyll content depending upon the extent of salt tolerance. Various studies have shown that high salt content results in stomatal closure, restricting the CO_2 fixation, reducing the chlorophyll content which in turn decreases the rate of photosynthesis in plants (El Sabagh et al., 2021; Irshad et al., 2022; Alghabari and Shah, 2021). The humic substances are plant biostimulants and regulate plant physiological processes for sustainable growth particularly under stress (Canellas and Olivares, 2014). Humic acid (HA) protects plant photosynthetic apparatus under various abiotic stresses and improves the chlorophyll and Pn as reported by Shukry et al. (2023). Similarly present study reported an improvement in chlorophyll and Pn in salt stressed wheat genotypes in the presence of HA, however this improvement varied based upon the type of wheat genotype i.e less improvement in susceptible BW, while high improvement in diverse SH (Figure 1). Alteration in osmolytic content under salt stress is portrayed as a plant strategical tool to counter the hazards of salt stress (Shah et al., 2017). However, application of HA decreases the concentration of proline, GB and Na^+ in salt stressed plant based upon their stress tolerance tendency as explicated by Khedr et al. (2022); Shah et al. (2018) and Hamed (2021) respectively. Likewise, Meng et al. (2023) concluded that reduction in proline content of salt stressed plants under HA application confirms the role of HA as a stress reliever, suggesting that plants get rid from osmotic stress, thereby minimizing the need of proline as an osmoprotectant. Likewise, according to Carillo et al. (2008) the HA treatment reduced the GB content salt stressed wheat plants through improving the water status and homeostasis, hence minimizing the need of GB deposition. On the other hand, salt stress induces different types of oxidative stresses within cellular system owe to the production of ROS (Alghabari and Shah, 2021). In this context, HA triggers the plant antioxidant defense system and increases the catalytic activity of antioxidant enzymes to relieve the plant from the damages caused by oxidative stress (Abbas et al., 2022). Correspondingly, in current research the HA reduced the GB and proline content in all salt stressed wheat genotypes. Besides, the SH wheat genotypes illustrated more dramatic reduction in proline and GB as compared to susceptible BW cultivars (Figure 1). In the same way, HA enhanced the activity of antioxidant enzymes CAT, POD and SOD in all salt stressed genotypes with high enhancement in SH as compare to BW susceptible genotypes (Figure 2).

In present study the HA treatment has reduced the accumulation of osmoprotectants such as proline and GB in salt stressed wheat plants, primarily owe to its role in stress mitigation and physiological stability. Plants usually synthesize these osmolytes under salinity stress in order to alleviate osmotic imbalance and oxidative stress. As HA improves water retention, plant water uptake and soil structure, therefore reducing the need for excessive accumulation of osmolytes (Canellas et al., 2015). Besides, HA triggers the activity of antioxidant enzymes such as SOD, POD and CAT, which significantly reduces the accumulation of ROS, hence declining the stress induced production of GB and proline (Aydin et al., 2012). Additionally, HA promotes ion homeostasis and nutrient uptake, specifically by increasing Na/K selectivity, hence alleviating the ionic toxicity and reducing the necessity for osmoprotectant based stress adaptation (Nardi et al., 2016). Besides, HA treatment also promotes the production of cytokinins, auxins and gibberellins, which alters plant metabolic

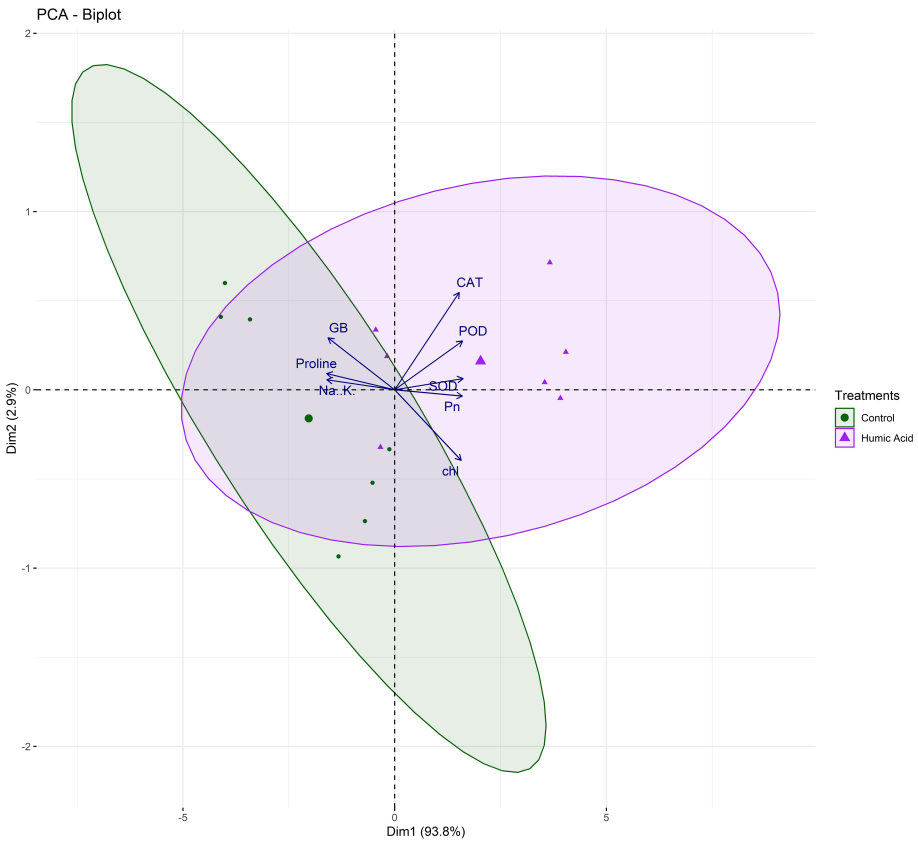


FIGURE 5
The varying dispersion of ellipses in biplot indicates the strong impact of HA treatment on chlorophyll, photosynthesis and antioxidant enzymes (SOD, POD and CAT), and strong impact of control treatment on osmolytes (Proline, andGB) and Na^+/K^+ in salt stressed wheat genotypes. Chl, chlorophyll; Pn, photosynthesis; GB, glycine betaine; SOD, superoxide dismutase; POD, peroxidase; CAT, catalase.

mode from stress tolerance to growth and development, resulting to a fall is osmolytes deposition (Zhang et al., 2020). Furthermore, HA triggers the chlorophyll and Pn, that further promote plant health and decreasing the plant dependence on stress-induced proline and GB production (Karakurt et al., 2009). As a whole, the HA tendency to counteract osmotic-stress, increase ion-homeostasis, boost antioxidant defense, and induce growth metabolism, explicates the recorded decline in proline and GB concentrations in salt stressed wheat plants.

In past, many studies recorded the potential role of HA in accelerating the vacuolar compartmentalization of Na^+ and amplifying the influx of K^+ that reduces the Na^+/K^+ and lowered the risk of sodium toxicity under salinity (Mona et al., 2017; Hamed, 2021; Abbas et al., 2022). Complementary, in present study HA lowered the Na^+/K^+ ratio in all salt stressed wheat SH and BW genotypes, with maximum reduction in SH as compared to susceptible BW genotypes (Figure 2). The physiological and biochemical processes are interlinked and vary in unison when plants are exposed to abiotic stresses (Shah et al., 2017). These processes as a whole determine the plant response to the stress. The plant biostimulants improve the response of plants against stress through strengthening the association of physiological and biochemical indicators of stress as confirmed by the present study (Figure 3). Furthermore, Alsamadany (2022) recorded unanimous decline in proline, GB and Na^+/K^+ content, while a unanimous rise in the antioxidant activity of enzymes (SOD, POD and CAT), Pn and Chl in salt stressed genotypes due to the application of HA. Correspondingly, current study rectified these findings in salt stressed SH and BW genotypes due to the application of HA (Figure 4 and 5). Besides, the salt tolerant SH genotypes had shown the strong association among chl, Pn and the catalytic activity of antioxidant enzymes, while the salt susceptible BW genotype illustrated the strong association among proline, GB and Na^+/K^+ratio (Figure 6) as reported by Alghabari and Shah (2024). This may be attributed to the role of humic substances in mediating the crosstalk between physiological and biochemical indices of stress tolerance as reviewed by Shah et al. (2024). Moreover, the results from PCA and heatmap further endorsed the potential role of HA in mitigating the effect of salt stress, and high tolerance of SH genotypes to the salt stress based upon the expression of osmolytic, antioxidant and physiological traits (Figures 4–7). Plants show intraspecific variation in tolerance to abiotic stresses, and response to biostimulants owe their different genetic texture (Shah et al., 2018). In this context, it is important to correlate the physio-chemical mechanisms of stress tolerance to genetic regulators. For instance, the proline producing gene *TaP5CS* overexpress in all salt stressed wheat genotypes to protect the subcellular structure from oxidative stress as reported by

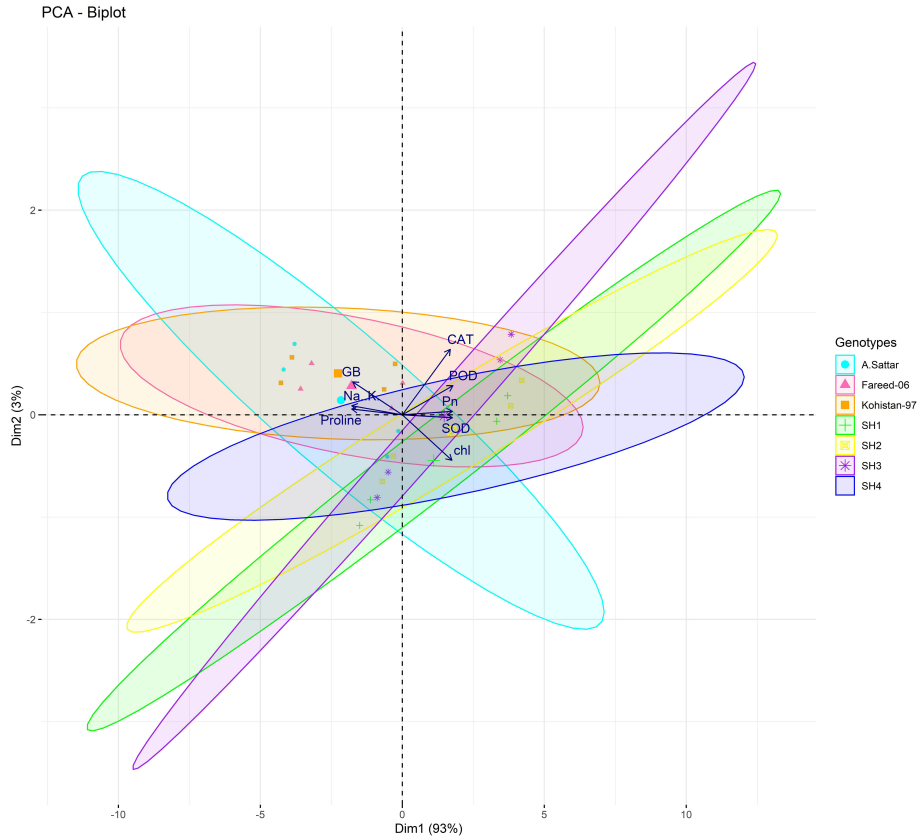


FIGURE 6
The varying dispersion of ellipses representing salt tolerant SHs (SH1-SH4) and salt susceptible BW in biplot, indicates the high expression of chlorophyll, photosynthesis and antioxidant enzymes (SOD, POD and CAT) in SH, and high expression of osmolytes (Proline, GB) and Na/K in BW genotypes. Chl, chlorophyll; Pn, photosynthesis; GB, glycine betaine; SOD, superoxide dismutase; POD, peroxidase; CAT, catalase.

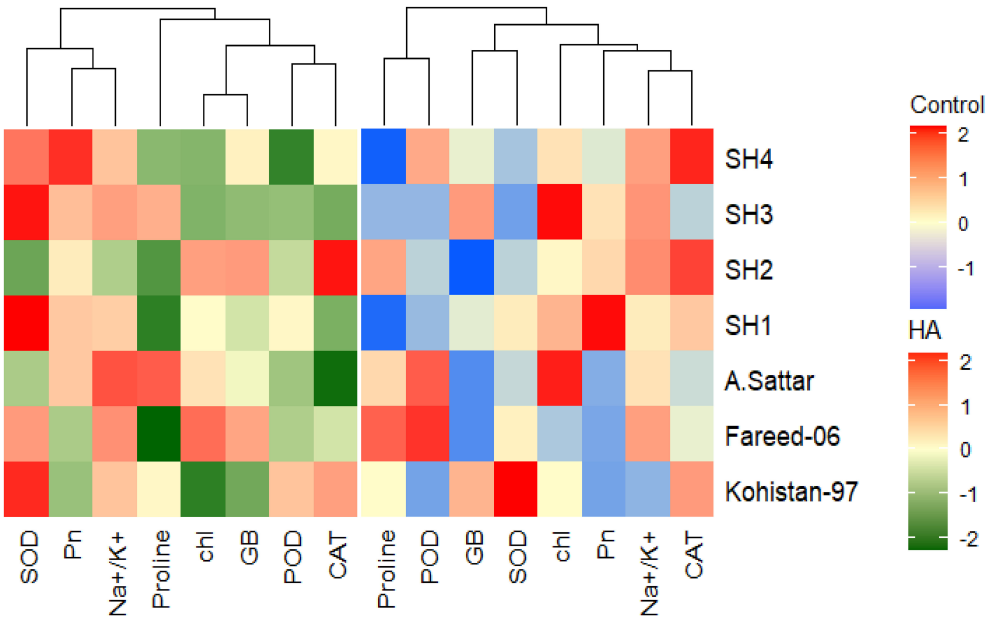


FIGURE 7
Heatmap categorizing the salt stressed wheat genotypes in terms of the varied trait expression under humic acid and control treatment. The varying colour pattern of bands from light to dark represent the change in the expression of trait from low to high due to the application of HA and control treatment. Chl, chlorophyll; Pn, photosynthesis; GB, glycine betaine; SOD, superoxide dismutase; POD, peroxidase; CAT, catalase.

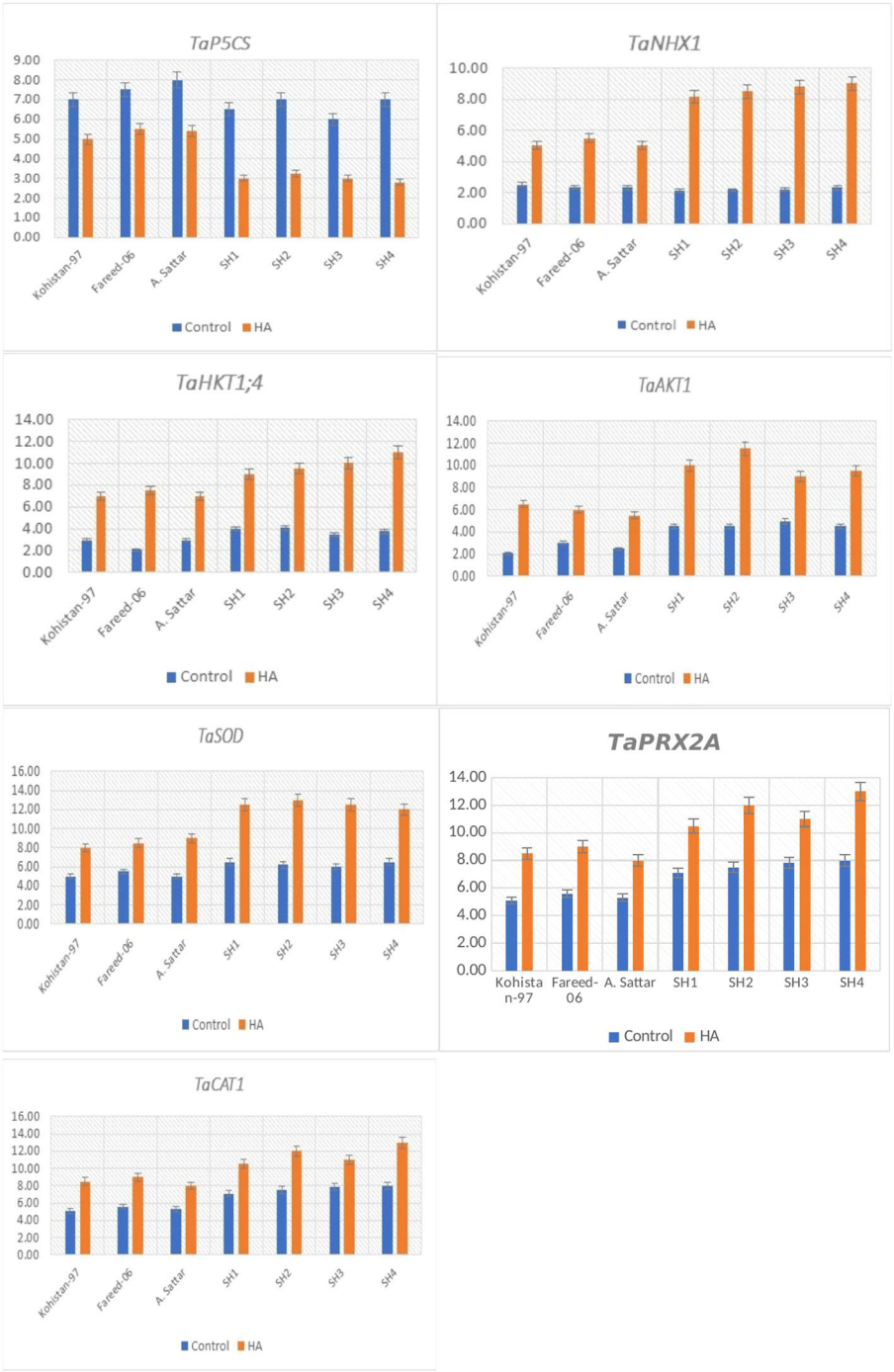


FIGURE 8
Relative expression analysis of salt stress related genes in susceptible BW and SH wheat genotypes under the application of HA and control treatment.

Alghabari and Shah (2024). Conversely, Meng et al. (2023) recorded decline in relative expression of *P5CS* gene in salt stressed *Lolium perenne* L due to HA. In parallel, current study reported downregulation of *TaP5CS* along with declined production of proline in all salt stressed wheat genotypes due to supplementation of HA (Figure 8). Corresponding to previous results this downregulation was more prominent in SH (SH1 to SH4) genotypes as compared to salt susceptible BW (Kohistan-97, Fareed-06 and A. Sattar) genotypes. Various studies in past rectified the potential role of HA in enhancing the activities of SOD, POD and CAT to reverse the effects of oxidative stress produced by salinity (Abbas et al., 2022; Meng et al., 2023). This is attributed to signaling role of HA that triggers the expression of *TaPOD*, *TaSOD* and *TaCAT* as endorsed by the present study (Figure 8). According to Al-Ashkar et al. (2019) less cellular Na^+/K^+ and detoxification of ROS are two main elements determining

plants tolerance to salt stress. Hence, the antioxidant proteins (SOD, POD, CAT) are essential for ROS scavenging, while the transporter proteins *HKT1* and *NHX1* are mandatory for Na^+ compartmentalization (El-Hendawy et al., 2017). Besides, Zeeshan et al. (2020) and, Alghabari and Shah (2024) explained the role of overexpressing inward-rectifying K^+ channel, *TaAKT1*, in speedy influx of K^+ in salt stressed wheat genotypes to lower the Na^+/K^+ ratio. Interestingly, in present study qRT-PCR analysis recorded high relative expression of *TaNHX1* and *TaHKT1,4*, *TaAKT1* along with increased influx of K^+ in all salt stressed wheat genotypes due to HA, with high expression in SH (SH1 to SH4) as compared to BW (Kohistan-97, Fareed-06 and A. Sattar) susceptible genotypes. In fact, HA enhances the protein abundance of *HKT1, NHX1* and *AKT1* in root stele that triggers the reabsorption of Na^+ from xylem vessel into neighboring cells, consequently less Na^+ is translocated to shoot and leaves (Khaleda et al., 2017).

Overall, in present study the HA triggered the physio-chemical and genetic indicators of salinity tolerance in all salt stressed wheat genotypes (Figure 9). In addition to role as stress reliever, the scalability and availability of HA for agricultural purposes is significantly important. HA is obtained from naturally occurring humic substances present in lignite, peat, and composted organic matter, making it globally and extensively available (Abbas et al., 2022). However, broad-scale application in progressive and commercial farming requires cost effective production and standardized extraction to ensure agricultural efficacy and sustainability (Shah et al., 2018). HA can significantly increase the soil fertility and reduce the dependency on chemical inputs; however, it cannot completely replace the synthetic fertilizers (Tahir et al., 2022). Although HA helps the plants to develop better tolerance and resilience to environmental stress, but it alone unable to provide sufficient nitrogen, phosphorus and potassium for plant growth. Therefore, the best approach is an integrated nutrient management (INM), mixing HA with organic and synthetic fertilizers, which ensure optimal plant growth and yield. Generally, the SH (SH1 to SH4) genotypes manifested high sensitivity to HA and showed high tolerance to salt stress as compared to susceptible BW (Kohistan-97, Fareed-06 and A. Sattar) genotypes based upon physio-chemical and genetic evaluations. Synthetic hexaploid (SH) wheat has high salt tolerance as compared to domesticated BW owe to wide genetic base inherited from its wild progenitors and less genetic erosion. In fact, the efficient response of SH wheat to salt stress is associated with rich genetic diversity inherited from its least explored D genome as reported by Yang et al. (2014).

5 Conclusion

Conclusively, the SH is bridge to transfer alien traits associated with salt tolerance from wild progenitors into elite bread wheat germplasm. The germplasm comparatively evaluated in current study will set a footprint for devising future breeding strategy to impart existing wheat cultivars the tolerance against salt stress.

While present study demonstrated the beneficial effects of HA in mitigating the salt stress under control conditions, there is a dire need for field level validation under varying climatic and soil conditions. The efficiency of HA in the soil of different regions based on soil types, temperature, pH, irrigation practices and organic matter content. Therefore, in future field level trials across different zones are mandatory to test the efficacy of HA in optimizing the wheat production. Furthermore, the large-scale evaluation will provide pragmatic insights into the long-term effects of HA on salt tolerance mechanisms of wheat.

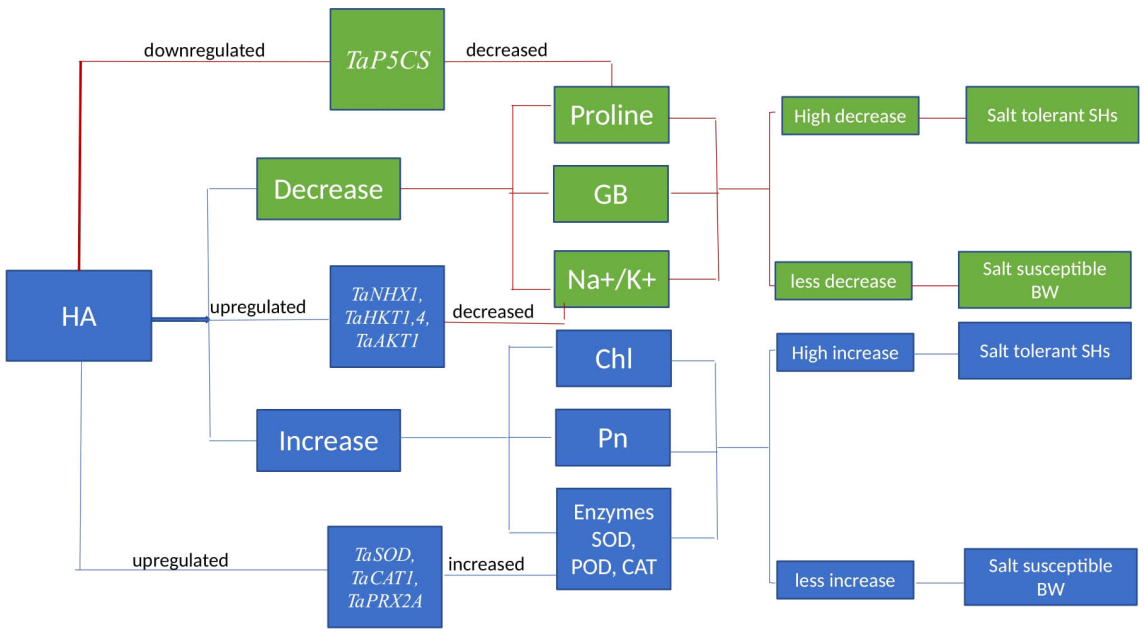


FIGURE 9
Schematic representation indicating the impact of HA on physiological, biochemical and molecular mechanisms imparting salt stress tolerance. Chl, chlorophyll; Pn, photosynthesis; GB, glycine betaine; SOD, superoxide dismutase; POD, peroxidase; CAT, catalase.

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

ZS: Writing – review & editing. FA: Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. Institutional Fund Projects under grant no. (IFPIP-692–155-1443) by the Ministry of Education and King Abdulaziz University, DSR, Jeddah, Saudi Arabia.

Acknowledgments

This research work was funded by the institutional fund project under grand number (IFPIP: 692-155-1443). The authors gratefully acknowledge the financial and technical support provided by Ministry of Education and King Abdulaziz University, DSR, Jeddah, Saudi Arabia.

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

References

Abbas, G., Rehman, S., Siddiqui, M. H., Ali, H. M., Farooq, M. A., and Chen, Y. (2022). Potassium and humic acid synergistically increase salt tolerance and nutrient uptake in contrasting wheat genotypes through ionic homeostasis and activation of antioxidant enzymes. *Plants* 11, 263. doi: 10.3390/plants11030263

Ahmad, A., Blasco, B., and Martos, V. (2022). Combating salinity through natural plant extracts based biostimulants: a review. *Front. Plant Sci.* 13, 862034. doi: 10.3389/fpls.2022.862034

Ahmad, I., Mian, A., and Maathuis, F. J. (2016). Overexpression of the rice AKT1 potassium channel affects potassium nutrition and rice drought tolerance. *J. Experim. Bot.* 67, 2689–2698. doi: 10.1093/jxb/erw103

Ahmed, K., Shabbir, G., Ahmed, M., Noor, S., Din, A. M. U., Qamar, M., et al. (2022). Expression profiling of TaARGOS homoeologous drought responsive genes in bread wheat. *Sci. Rep.* 12, 3595. doi: 10.1038/s41598-022-07637-y

Al-Ashkar, I., Alderfasi, A., El-Hendawy, S., Al-Suhaibani, N., El-Kafafi, S., and Seleiman, M. F. (2019). Detecting salt tolerance in doubled haploid wheat lines. *Agronomy* 9, 211. doi: 10.3390/agronomy9040211

Alghabari, F., and Shah, Z. H. (2024). Comparative adaptability assessment of bread wheat and synthetic hexaploid genotypes under saline conditions using physiological, biochemical, and genetic indices. *Front. Plant Sci.* 15, 1336571. doi: 10.3389/fpls.2024.1336571

Ali, S., Rizwan, M., Qayyum, M. F., Ok, Y. S., Ibrahim, M., Riaz, M., et al. (2022). Potassium and humic acid synergistically increase salt tolerance in wheat by improving plant growth, antioxidant defense system, and ionic homeostasis. *Front. Plant Sci.* 13, 8840195.

Alici, E. H., and Arabaci, G. (2016). Determination of SOD, POD, PPO and cat enzyme activities in Rumex obtusifolius L. *Ann. Res. Rev. Biol.* 11, 1–7. doi: 10.9734/ARRB/2016/29809

Alsamadany, H. (2022). Physiological, biochemical and molecular evaluation of mungbean genotypes for agronomical yield under drought and salinity stresses in the presence of humic acid. *Saudi J. Biol. Sci.* 29, 103385. doi: 10.1016/j.sjbs.2022.103385

Alsamadany, H., Alzahrani, Y., and Shah, Z. H. (2023). Physiomorphic and molecular-based evaluation of wheat germplasm under drought and heat stress. *Front. Plant Sci.* 14, 1107945. doi: 10.3389/fpls.2023.1107945

Aycan, M., Baslam, M., Mitsui, T., and Yildiz, M. (2022). The TaGSK1, TaSRG, TaPTF1, and TaP5CS gene transcripts confirm salinity tolerance by increasing proline production in wheat (Triticum aestivum L.). *Plants* 11, 3401. doi: 10.3390/plants11233401

Aydin, A., Kant, C., and Turan, M. (2012). Humic acid application alleviates salt stress and improves growth parameters in wheat (Triticum aestivum L.). *J. Plant Nutri.* 35, 567–580.

Canellas, L. P., and Olivares, F. L. (2014). Physiological responses to humic substances as plant growth promoter. *Chem. Biol. Tech. Agric.* 1, 3. doi: 10.1186/2196-5641-1-3

Canellas, L. P., Olivares, F. L., Aguiar, N. O., Jones, D. L., Nebbioso, A., Mazzei, P., et al. (2015). Humic and fulvic acids as biostimulants in agriculture. *J. Plant Physiol.* 206, 36–44.

Carillo, P., and Gibon, Y. (2011). Protocol: Extraction and determination of proline. *PrometheusWiki*. 1–5.

Carillo, P., Mastrolonardo, G., Nacca, F., Parisi, D., Verlotta, A., and Fuggi, A. (2008). Nitrogen metabolism in durum wheat under salinity: accumulation of proline and glycine betaine. *Func. Plant Biol.* 35, 412–426. doi: 10.1071/FP08108

Chaudhary, D., Pal, N., Arora, A., Prashant, B. D., and Venadan, S. (2024). "Plant functional traits in crop breeding: advancement and challenges," in *Plant Functional Traits for Improving Productivity* (Springer Nature Singapore, Singapore), 169–202.

Dave, A., Agarwal, P., and Agarwal, P. K. (2022). Mechanism of high affinity potassium transporter (HKT) towards improved crop productivity in saline agricultural lands. *3 Biotech.* 12, 51. doi: 10.1007/s13205-021-03092-0

El-Hendawy, S., Hassan, W. M., Al-Suhaibani, N. A., Refay, Y., and Abdella, K. A. (2017). Comparative performance of multivariable agro-physiological parameters for detecting salt tolerance of wheat cultivars under simulated saline field growing conditions. *Front. Plant Sci.* 8, 435. doi: 10.3389/fpls.2017.00435

El Sabagh, A., Islam, M. S., Skalicky, M., Ali Raza, M., Singh, K., Anwar Hossain, M., et al. (2021). Salinity stress in wheat (Triticum aestivum L.) in the changing climate: Adaptation and management strategies. *Front. Agron.* 3, 661932.

FAO (2024) *FAO launches first major global assessment of salt-affected soils in 50 years*. Available at: https://www.fao.org.

Grieve, C. M., and Grattan, S. R. (1983). Rapid assay for determination of water-soluble quaternary ammonium compounds. *Plant Soil* 70, 303–307. doi: 10.1007/BF02374789

Gupta, B., and Huang, B. (2014). Mechanism of salinity tolerance in plants: physiological, biochemical, and molecular characterization. *Intern. J. Genomic.* 1, 701596. doi: 10.1155/2014/701596

Hamed, E. N. (2021). Effect of salicylic, humic and fulvic acids application on the growth, productivity and elements contents of two wheat varieties grown under salt stress. *J. Soil Sci. Agric. Engin* 12, 657–671.

Hasanuzzaman, M., Raihan, M. R. H., Masud, A. A. C., Rahman, K., Nowroz, F., Rahman, M., et al. (2021). Regulation of reactive oxygen species and antioxidant defense in plants under salinity. *Intern. J. Mol. Sci.* 22, 9326. doi: 10.3390/ijms22179326

Havre, G. N. (1961). The flame photometric determination of sodium, potassium and calcium in plant extracts with special reference to interference effects. *Analytica Chim. Acta* 25, 557–566. doi: 10.1016/S0003-2670(01)81614-7

Hussain, S., Hussain, S., Ali, B., Ren, X., Chen, X., Li, Q., et al. (2021). Recent progress in understanding salinity tolerance in plants: Story of Na+/K+ balance and beyond. *Plant Physiol. Biochem.* 160, 239–256. doi: 10.1016/j.plaphy.2021.01.029

Irshad, A., Ahmed, R. I., Ur Rehman, S., Sun, G., Ahmad, F., Sher, M. A., et al. (2022). Characterization of salt tolerant wheat genotypes by using morpho-physiological, biochemical, and molecular analysis. *Front. Plant Sci.* 13, 956298. doi: 10.3389/fpls.2022.956298

Jiang, W., Yang, L., He, Y., Zhang, H., Li, W., Chen, H., et al. (2019). Genome-wide identification and transcriptional expression analysis of superoxide dismutase (SOD) family in wheat (*Triticum aestivum*). *Peer J.* 19, 7e8062. doi: 10.7717/peerj.8062

Karakurt, H., Unlu, H., and Padem, H. (2009). The influence of humic acid on nutrient content and uptake in tomato (Lycopersicon esculentum L.) under salinity stress. *J. Plant Nutr.* 32, 20–30.

Khaled, H., and Fawy, H. A. (2011). Effect of different levels of humic acids on the nutrient content, plant growth, and soil properties under conditions of salinity. *Soil Water Res.* 6 (1), 21–29.

Khaleda, L., Park, H. J., Yun, D. J., Jeon, J. R., Kim, M. G., Cha, J. Y., et al. (2017). Humic acid confers high-affinity K+ transporter 1-mediated salinity stress tolerance in Arabidopsis. *Molecules Cells* 40, 966–975.

Khedr, R. A., Sorour, S. G. R., Aboukhadrah, S. H., El Shafey, N. M., Abd Elsalam, H. E., El-Sharnouby, M. E., et al. (2022). Alleviation of salinity stress effects on agro-physiological traits of wheat by auxin, glycine betaine, and soil additives. *Saudi J. Biol. Sci.* 29, 534–540. doi: 10.1016/j.sjbs.2021.09.027

Li, A., Liu, D., Yang, W., Kishii, M., and Mao, L. (2018). Synthetic hexaploid wheat: yesterday, today, and tomorrow. *Engineering* 4, 552–558. doi: 10.1016/j.eng.2018.07.001

Mcgraw-Hill, C. (2008). *Statistix 8.1 (Analytical software, tallahassee, florida)* (Florida, USA: Maurice/Thomas text).

Mehla, N., Sindhi, V., Josula, D., Bisht, P., and Wani, S. H. (2017). An introduction to antioxidants and their roles in plant stress tolerance. *Reactive Oxygen Species Antioxid. Syst. Plants: Role Regul. Under Abiotic Stress*, 1–23.

Meng, Q., Yan, M., Zhang, J., Zhang, Q., Zhang, X., Yang, Z., et al. (2023). Humic acids enhance salt stress tolerance associated with pyrroline 5-carboxylate synthetase gene expression and hormonal alteration in perennial ryegrass (Lolium perenne L.). *Front. Plant Sci.* 14, 1272987. doi: 10.3389/fpls.2023.1272987

Mona, I. N., Gawish, S. M., Taha, T., and Mubarak, M. (2017). Response of wheat plants to application of selenium and humic acid under salt stress conditions. *Egypt. J. Soil Sci.* 57, 175–187. doi: 10.21608/ejss.2017.3715

Nardi, S., Pizzeghello, D., Schiavon, M., and Ertani, A. (2016). Plant biostimulants: Physiological responses induced by protein hydrolyzed-based products and humic substances in plant metabolism. *Sci. Agricola* 73, 18–23. doi: 10.1590/0103-9016-2015-0006

RStudio Team. (2020). *RStudio: Integrated development for r* (Boston, MA: RStudio, PBC). Available at: http://www.rstudio.com/.

Sachdev, S., Ansari, S. A., Ansari, M. I., Fujita, M., and Hasanuzzaman, M. (2021). Abiotic stress and reactive oxygen species: Generation, signaling, and defense mechanisms. *Antioxidants* 10, 277. doi: 10.3390/antiox10020277

Shah, Z. H., Rehman, H. M., Akhtar, T., Alsamadany, H., Hamooh, B. T., Mujtaba, T., et al. (2018). Humic substances: Determining potential molecular regulatory processes in plants. *Front. Plant Sci.* 9, 263. doi: 10.3389/fpls.2018.00263

Shah, Z. H., Rehman, H. M., Akhtar, T., Daur, I., Nawaz, M. A., Ahmad, M. Q., et al. (2017). Redox and ionic homeostasis regulations against oxidative, salinity and drought stress in wheat (a systems biology approach). *Front. Genet.* 8, 41. doi: 10.3389/fgene.2017.00141

Shukry, W. M., Abu-Ria, M. E., Abo-Hamed, S. A., Anis, G. B., and Ibraheem, F. (2023). The efficiency of humic acid for improving salinity tolerance in salt sensitive rice (Oryza sativa): growth responses and physiological mechanisms. *Gesunde Pflanzen* 75, 2639–2653. doi: 10.1007/s10343-023-00885-6

Su, P., Yan, J., Li, W., Wang, L., Zhao, J., Ma, X., et al. (2020). A member of wheat class III peroxidase gene family, TaPRX-2A, enhanced the tolerance of salt stress. *BMC Plant Biol.* 20, 1–15. doi: 10.1186/s12870-020-02602-1

Tahir, M., Khurshid, M., Khan, M. Z., Abbasi, M. K., and Kazmi, M. H. (2022). Lignite-derived humic acid effect on growth of wheat plants in different soils. *Pedosphere* 22, 817–824.

Teleubay, Z., Yermekov, F., Rustembayev, A., Topayev, S., Zhabayev, A., Tokbergenov, I., et al. (2024). Comparison of climate change effects on wheat production under different representative concentration pathway scenarios in North Kazakhstan. *Sustainability* 16, 293.

Yang, C., Zhao, L., Zhang, H., Yang, Z., Wang, H., Wen, S., et al. (2014). Evolution of physiological responses to salt stress in hexaploid wheat. *PNAS* 111, 11882–11887. doi: 10.1073/pnas.1412839111

Zeeshan, M., Lu, M., Naz, S., Sehar, S., Cao, F., and Wu, F. (2020). Resemblance and difference of seedling metabolic and transporter gene expression in high tolerance wheat and barley cultivars in response to salinity stress. *Plants* 9, 519. doi: 10.3390/plants9040519

Zhang, J., Meng, Q., Yang, Z., Zhang, Q., Yan, M., Hou, X., et al. (2024). Humic acid promotes the growth of switchgrass under salt stress by improving photosynthetic function. *Agronomy* 14, 1079. doi: 10.3390/agronomy14051079

Zhang, X., Wang, K., and Fang, C. (2020). Humic acid alleviates salt stress in wheat seedlings by improving plant growth, photosynthetic efficiency, and osmotic regulation. *Plant Physiol. Biochem.* 151, 53–63.

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 16 May 2025
DOI 10.3389/fpls.2025.1583953

OPEN ACCESS

EDITED BY
Diaa Abd El Moneim,
Arish University, Egypt

REVIEWED BY
Shujuan Zhao,
Shanghai University of Traditional Chinese Medicine, China
Thijs Van Gerrewey,
Ghent University, Belgium
Linhe Su,
Far Eastern Branch (RAS), Russia
Xiaofen Qi,
Heilongjiang University, China

*CORRESPONDENCE
Ann Abozeid
annabozeid@science.menofia.edu.eg
Zongqi Yang
yangzongqi@zstu.edu.cn
Dongfeng Yang
yangdongfeng@zstu.edu.cn

RECEIVED 11 March 2025
ACCEPTED 18 April 2025
PUBLISHED 16 May 2025

CITATION
Abozeid A, Du X, Zhang L, Yang F, Wu J, Zhang L, Cui Q, Yang Z and Yang D (2025) Comparative transcriptomic analysis and genome-wide identification provide insights into the potential role of fungal-responsive MAPK cascade genes in tanshinone accumulation in *Salvia miltiorrohiza*. *Front. Plant Sci.* 16:1583953. doi: 10.3389/fpls.2025.1583953

Comparative transcriptomic analysis and genome-wide identification provide insights into the potential role of fungal-responsive MAPK cascade genes in tanshinone accumulation in *Salvia miltiorrohiza*

Ann Abozeid[1,2]*, Xinru Du[1], Lan Zhang[1], Furui Yang[1], Jianxiong Wu[3], Lin Zhang[1], Qi Cui[4], Zongqi Yang[1]* and Dongfeng Yang[1,5]*

[1]College of Life Sciences and Medicine, Key Laboratory of Plant Secondary Metabolism and Regulation of Zhejiang Province, Zhejiang Sci-Tech University, Hangzhou, China, [2]Botany and Microbiology Department, Faculty of Science, Menoufia University, Shebin Elkoom, Egypt, [3]College of Plant Protection, China Agricultural University, Beijing, China, [4]Laboratory of Ornamental Plants, Department of Landscape Architecture, Zhejiang Sci-Tech University, Hangzhou, China, [5]Shaoxing Biomedical Research Institute of Zhejiang Sci-Tech University Co., Ltd, Zhejiang Engineering Research Center for the Development Technology of Medicinal and Edible Homologous Health Food, Shaoxing, China

Salvia miltiorrhiza is a well-known traditional Chinese medicine (TCM) for its bioactive tanshinones that are used to treat various diseases and have high antimicrobial properties. Previous studies have shown that tanshinone accumulation in *S. miltiorrhiza* was shown to be significantly induced by fungal elicitors. Mitogen-activated protein kinases (MAPKs), which play critical roles in plant–microbe interactions and cellular processes, are known to regulate the accumulation of antimicrobial metabolites. In this study, we aimed to identify MAPK families in *S. miltiorrhiza* and screen SmMAPKs for candidates involved in fungal elicitor-mediated tanshinone accumulation. Through genome-wide analysis, we identified 17 MAPK, 7 MAPKK, and 22 MAPKKK genes in *S. miltiorrhiza*, which were distributed across nine chromosomes. Phylogenetic analysis classified SmMAPKs into two subgroups, TEY and TDY, similar to *Arabidopsis* MAPKs, while all SmMAPKKKs clustered under the MEKK subfamily. Cis-acting element analysis revealed that most SmMAPK genes are associated with stress and phytohormone responses suggesting their involvement in defense mechanisms. To investigate the role of MAPK s in tanshinone accumulation, hairy roots of *S. miltiorrhiza* were treated with two fungal elicitors, yeast extract and *Aspergillus niger*, for 1 and 4 days. HPLC analysis demonstrated that both elicitors significantly promoted the accumulation of tanshinones, particularly cryptotanshinone and dihydrotanshinone. Comprehensive transcriptomic analysis, followed by Pearson correlation coefficient analysis, revealed a strong positive correlation between tanshinone content and *SmMPK4* and *SmMPKK5*, while negative correlations were observed with *SmMPKKK6*, *SmMPKKK11*, and *SmMPKKK20*. The presence of defense-

related cis-acting elements in the promoter regions of *SmMPK4*, *SmMPKK5*, *SmMPKKK6*, *SmMPKKK11*, and *SmMPKKK20* further supports their involvement in fungal elicitor-mediated tanshinone accumulation. This study provides critical insights into the regulatory roles of SmMAPK genes in tanshinone accumulation in *S. miltiorrhiza* in response to fungal elicitors. These findings have potential applications in enhancing tanshinone production for medicinal purposes offering a foundation for further research into the molecular mechanisms underlying tanshinone biosynthesis.

KEYWORDS

***S. miltiorrhiza*, tanshinones, MAPK, fungal elicitors, phylogenetic analysis**

Introduction

Salvia miltiorrhiza is a well-known traditional Chinese medicine (TCM) valued for its production of tanshinones, the primary bioactive compounds (Xia et al., 2023; Huang et al., 2024). These diterpenoids are clinically used to treat hepatocirrhosis, cardiovascular diseases (Ren et al., 2019), and menstrual pain (Lee et al., 2020). Beyond their therapeutic applications, tanshinones exhibit antimicrobial properties, with cryptotanshinone and dihydrotanshinone I demonstrating stronger activity than tanshinone I and tanshinone IIA (Zhao et al., 2011), suggesting their role as defensive metabolites. Notably, fungal elicitors significantly enhance tanshinone accumulation in *S. miltiorrhiza* (Ming et al., 2013; Contreras et al., 2019) further supporting their involvement in plant defense. Transcriptomic studies reveal that fungal induction upregulates genes in the tanshinone biosynthetic pathway (Zhou et al., 2017; Wu et al., 2022), though the precise regulatory mechanisms remain unclear. Given their dual antimicrobial and pharmacological potential, elucidating the biosynthetic regulation of tanshinones could enable biotechnological strategies to optimize their production for medical and agricultural applications. Fungal elicitation starts with a pathogen-associated molecular pattern (PAMP) that binds to the plant cell membrane through recognition receptors. The binding of PAMPs and membrane receptors activates the mitogen-activated protein kinase (MAPK) signaling pathway that results in the alleviation of transcription factor (TF) expression, which, in turn, activates the defensive metabolites biosynthesis genes (Zhai et al., 2017). For example, defensive metabolites biosynthesis genes were activated by different types of transcription factors in corn (Ibraheem et al., 2015), cotton (Xu et al., 2004), and rice (Yamamura et al., 2015).

Mitogen-activated protein kinases (MAPKs) are protein kinases that are involved in signaling pathways of various cellular processes, such as growth and development, metabolism, cell death, and defense responses (Zhang and Liu, 2002; Roux and Blenis, 2004; Wang et al., 2007; Widmann et al., 1999; Khan et al., 2024). In addition, they play a crucial role in plant–microbe interactions and trigger immune responses (Arthur and Ley, 2013; Guo et al., 2021; Van Gerrewey and Chung, 2024). In addition to their role in stress responses, MAPK cascades have been shown to regulate the biosynthesis of secondary metabolites in various plant species, including camalexin in *Arabidopsis* and momilactones in rice (Ren et al., 2008; Kishi-Kaboshi et al., 2010a; Kishi-Kaboshi et al., 2010b; Gaur et al., 2018; Li et al., 2024; Zhou et al., 2025). These findings highlight the potential of MAPK cascades as key regulators of defensive metabolite biosynthesis in plants. However, MAPKs have not been identified or characterized before in *S. miltiorrhiza*, and their role in fungal elicitor-mediated tanshinone accumulation has not been investigated yet.

In this study, we detected and comprehensively analyzed the SmMAPK gene families in *S. miltiorrhiza*. SmMAPK gene chromosomal distribution, phylogenetic relationships, protein alignments, conserved motifs, conserved domains, cis-acting elements, and gene structure were investigated. We treated hairy roots of *S. miltiorrhiza* with two fungal elicitors, yeast extract and *Aspergillus niger*, that significantly promoted tanshinone accumulation. Then, we employed Pearson correlation coefficient analysis to screen SmMAPK genes that may have a potential role in tanshinone accumulation in *S. miltiorrhiza*. This study provides new insights into revealing the potential roles of SmMAPK genes in regulating the fungal elicitor-mediated tanshinone accumulation in *S. miltiorrhiza*.

Material and methods

Plant materials and treatments

The hairy roots of *S. miltiorrhiza* were cultured at Zhejiang Sci-Tech University, Hangzhou, China. Approximately 0.2 g of hairy roots with the same chronological age (21 days post-induction) and active growth status (visible root tip elongation) to ensure biological consistency across replicates were inoculated into 50 ml of MS 6,7-V liquid culture medium. The hairy roots were cultured in a shaking incubator at 110 rpm/min. Three-week-old hairy roots were treated

with 100 mg/L of yeast extract and *A. niger* elicitors. After 1, 4, and 7 days, hairy roots were collected for HPLC and RNA sequencing. Immediately after harvest, hairy root samples were flash-frozen in liquid nitrogen, stored at −80°C in RNase-free tubes, and transferred to dry ice during transport to prevent thawing. Each group had three biological replicates.

Preparation of elicitors

Aspergillus niger ATCC 6275 freeze-dried powder was purchased from Hunan Fenghui Biotechnology Co., Ltd. and yeast extract powder (LP0021B) from Oxoid. The *Aspergillus niger* freeze-dried powder was activated; then, the elicitor was prepared using the acid hydrolysis method of Wen-Zhi et al. (2002), while we prepared the yeast extract elicitor following the method of Shi et al. (2014). Both elicitors' mass concentrations were expressed as sugar mass concentrations. The sugar concentration was determined using the anthrone colorimetric (micro-method) soluble sugar content determination kit. The glucose standard curve drawn by the glucose content (x) against the absorbance (y) in the kit is y = 4.275x − 0.07, where x and y represent standard concentration (mg/ml) and absorbance value, respectively.

Extraction and HPLC detection of tanshinones from hairy roots

Air-dried hairy roots for 4 days at 40°C were ground into powder using a sample grinder (Osheng Instrument, Bioprep-24). Into a centrifuge tube, 0.02 g was weighed using an analytical balance, and 1 ml of 70% methanol (Macklin, M813907-4L) solution was added. The tube was placed in an ultrasonic extractor for ultrasonic extraction for 1 h. The sample was then centrifuged at 12,000 rpm for 15 min, and the supernatant was filtered using a 0.22-μm oil filter (Biosharp, BS-QT-013). The filtrate was collected in a sample injection bottle. A phosphoric acid water with 0.02% concentration was prepared as the mobile phase, using ultrapure water and phosphoric acid (Macklin, P816338—500 ml) after filtration, and then ultrasonic degassing was performed. Secondary metabolites were detected by high-performance liquid chromatography under the following: high-performance liquid chromatography (Waters 2695) and diode array detector (Waters 2998) were used for detection, the chromatographic column was Waters SunFire C18 column, chromatographic acquisition and analysis were completed by Empower 2 software, and the loading volume was 20 μl. The chromatographic conditions used were a flow rate of 1 ml/min and a column temperature of 25°C. The detection wavelengths were 270 nm. External standard analyte peak area was used to determine the tanshinone compound content. Three replicates were used for all samples to ensure accuracy.

Extraction and sequencing of RNA

TRIzol® Reagent was used for RNA extraction. RNA integrity was evaluated with the 5300 Bioanalyzer (Agilent), and quantification was performed using the ND-2000 spectrophotometer (NanoDrop Technologies). Only high-quality RNA sample (OD260/280 = 1.8–2.2, OD260/230 ≥ 2.0, RIN ≥ 6.5, 28S:18S ≥ 1.0, > 1 μg) was used to construct a sequencing library.

RNA sequencing was done by Shanghai Majorbio Biopharm Biotechnology Co., Ltd., China, per the manufacturer's guidelines (Illumina, San Diego, CA). RNA-seq libraries were prepared using the Illumina® Stranded mRNA Prep, Ligation protocol. Shortly, messenger RNA was isolated according to the polyA selection method by oligo(dT) beads and then fragmented by fragmentation buffer first. Second, double-stranded cDNA was synthesized using a SuperScript double-stranded cDNA synthesis kit (Invitrogen, CA) with random hexamer primers (Illumina). Then, the synthesized cDNA was subjected to end repair, phosphorylation, and "A" base addition according to Illumina's library construction protocol. Libraries were size selected for cDNA target fragments of 300 bp on 2% Low Range Ultra Agarose followed by PCR amplified using Phusion DNA polymerase (NEB) for 15 PCR cycles. After quantification by Qubit 4.0, the paired-end RNA-seq sequencing library was sequenced with the NovaSeq 6000 sequencer (2 × 150-bp read length).

Identification and characterization of MAPK gene family in *S. miltiorrhiza*

The gene sequences of MAPKs of *Arabidopsis thaliana* (20 AtMAPKs, 10 AtMAPKKs, and 80 AtMAPKKKs) were downloaded from the *A. thaliana* database TAIR (Berardini et al., 2015), *Oryza sativa* (17 OsMAPKs, 8 OsMAPKKs, and 75 OsMAPKKKs) from the Rice Genome Annotation Project database (Kawahara et al., 2013), cucumber (14 CsMAPK, 6 CsMAPKK, and 59 CsMAPKKK) from the Cucumber Genomics Database, and the *S. miltiorrhiza* genome data were retrieved from the GWH database (Pan et al., 2023).

Nucleotide BLAST was carried out with a cutoff of E-value <1×10^{-5} as the query and search for the MAPK genes within the *S. miltiorrhiza* genome. HMM profiles of the protein kinase domain (PF00069) were sourced from the Pfam database, and the NCBI database (Wang et al., 2023) was used to confirm the presence of the conserved domain within the candidate MAPK.

The theoretical isoelectric point (pI) and molecular weight (MW) of the MAPK genes were determined using the Expasy online tool (Gasteiger et al., 2005). The subcellular localization of MAPK proteins was predicted via the BUSCA online platform (Savojardo et al., 2018). The chromosomal localization of MAPK genes was analyzed by the TBtools software (Chen et al., 2023).

Synteny analysis

To explore the evolutionary conservation of MAPK genes, we performed synteny analysis by comparing the genomic locations of MAPK genes in *S. miltiorrhiza* with those in closely related species such as *Arabidopsis thaliana* and *Oryza sativa*. We used MCScanX to identify orthologous genes and analyzed gene order conservation across species. Tbtools was used to visualize the synteny analysis maps.

Phylogenetic analysis, amino acid sequences alignments, conserved motifs, domain analyses, cis-acting regulatory elements, and gene structure analysis

Phylogenetic trees were generated using MEGA 11 software with the neighbor-joining (NJ) method, and a bootstrap analysis was conducted with 1,000 iterations. Phylogenetic trees were visualized using the iTOL v6 online tool (https://itol.embl.de/). Amino acid sequence alignments were done by MUSCLE, and GENEDOC software was used to visualize the alignments and detect the conserved motif signatures. Ten motifs were identified from SmMAPK proteins using the MEME online website (https://meme-suite.org/meme/tools/meme). Conserved domain analysis was done by the NCBI database (https://www.ncbi.nlm.nih.gov/). The cis-acting element analysis was done by PlantCARE online. Conserved domains, distribution of cis-acting elements in the promoters, and SmMAPK gene structure were visualized using TBtools software.

Protein–protein interaction network and functional annotation analysis

Protein–protein interaction (PPI) networks were constructed using the STRING database. Interactions were selected based on a confidence score threshold of 0.4. Gene Ontology (GO) enrichment analysis and KEGG pathway analysis were performed and visualized using the STRING database. Enrichment was evaluated using default parameters.

Differential expression analysis and Pearson correction analysis

Gene expression analysis was performed using RNA-seq data. Differential expression was determined with DESeq2 using thresholds of $|\log2FC| \geq 0.5$ and FDR < 0.05. Heatmaps visualizing normalized expression of differentially expressed genes (DEGs) were generated using TBtools. All fold change, log2 fold change values, regulation, and significance are summarized in Supplementary Table S5. For co-expression analysis, pairwise Pearson correlations were computed from normalized expression values and plotted as a correlation matrix using the SRplot online platform. Correlations with a p-value < 0.05 were considered statistically significant.

Gene expression validation by RT-qPCR

Real-time quantitative polymerase chain reaction was used to analyze the relative expression levels of SmMAPKs candidate genes. SYBR® Green Pro Taq HS premixed qPCR kit (AG, AG11701) was used for fluorescence quantitative PCR reaction. QuantStudio 6 Flex Real-Time PCR System (ThermoFisher, USA) was used to analyze the expression of SmMAPKs genes. Actin was chosen as the internal control for normalization in qPCR, and three biological replicates were used. The experiments were performed with the following operating parameters: sample pass initial denaturation at 95°C for 30 s, then pass 40 amplification cycles. Each cycle consisted of denaturation at 95°C for 5 s and annealing at 60°C for 30 s. The last stage of dissociation was 95°C for 15 s, 65°C for 1 min, and 95°C for 15 s. Primers used for RT-qPCR are listed in Supplementary Table S6.

Statistical analysis

Data analysis and ANOVA tests were done by Prism 9.0.0. Error bars represent the mean value of standard deviation ± for three independent replicas. One-way ANOVA multiple comparison test was used to determine significant differences. (****) represents a significant difference at the level of $p < 0.0001$ relative to the control group.

Results

Genome-wide identification and chromosomal mapping of MAPK genes in *S. miltiorrhiza*

A total of 17 MAPK, 7 MAPKK, and 22 MAPKKK genes were identified from *S. miltiorrhiza* genome (Supplementary Table S1). The sequence of cDNA and proteins of identified MAPK genes are listed in Supplementary Table S2 and Supplementary Table S3, respectively. The cDNA length of identified MAPK genes ranged from 708 bp in *SmMPKKK20* to 2,877 bp in *SmMPKKK2*. The predicted MW of the MAPK proteins ranged from 25,860.58 Da in *SmMPKKK20* to 95,481.2 Da in *SmMPKKK21*, while their theoretical pI ranged from 4.51 in *SmMPKKK6* to 9.52 in *SmMPKKK21*. Subcellular localization analysis revealed that 32 MAPK proteins are localized in the nucleus, 7 MAPK proteins were predicted to localize in the chloroplast, 6 proteins are located in the cytoplasm, while only *SmMPKKK22* is located in the

endomembrane system. Chromosome mapping revealed that *S. miltiorrhiza* MAPK genes are located on nine chromosomes (Figure 1). Chromosome 8 has the biggest number of MAPK genes as it has nine genes, while only one gene (*SmMPKKK7*) is located on chromosome 1. Chromosomes 5, 7, and 9 all have the same number of MAPK genes (seven) while chromosome 2 has five genes, and both chromosomes 3 and 5 have three genes.

Synteny analysis

Our synteny analysis (Figure 2) revealed that 21 MAPK genes in *S. miltiorrhiza* exhibit conservation with orthologs in *A. thaliana* and 4 genes in *O. sativa*. This suggests that the WAK/WAKL family has been preserved throughout evolution.

Phylogenetic analysis of MAPK genes

To investigate the evolutionary relationships of the identified MAPK genes, a phylogenetic tree was generated (17 MAPKs, 7 MAPKKs, and 22 MAPKKKs) from *S. miltiorrhiza*, together with 20 MAPKS, 10 MAPKKS, and 80 MAPKKKS from *Arabidopsis thaliana*, using the neighbor-joining (NJ) method. The protein sequences of *Arabidopsis* MAPK are listed in Supplementary Table S3. The MAPK proteins of each subfamily from the two species were grouped together in a separate group. Moreover, most of the MAPK proteins from *S. miltiorrhiza* clustered with their homologs in *A. thaliana*. Interestingly, all proteins from the family MAPKKKS clustered with *A. thaliana* MEKK subfamily (Figure 3).

MAPK characterization

The *Salvia miltiorrhiza* MAPK (SmMAPK) gene family was characterized through comparative analysis with *Arabidopsis thaliana*. Phylogenetic tree construction (Figure 4A) revealed that all 17 SmMAPK proteins cluster into the two conserved *Arabidopsis* subgroups defined by their TxY phosphorylation motifs (TEY or TDY) (Ichimura et al., 2002b). Sequence alignments confirmed these motifs in SmMAPKs (Figure 4B). MEME analysis identified 10 conserved motifs (Figure 4C; Supplementary Table S4), with motifs 2, 5, 8, and 9 forming the Pkinase domain. Motif 7 was exclusive to the TDY-subtype C-termini, while motif 10 predominated in the TEY-subtype members. Domain analysis (Figure 4D) showed that all SmMAPKs contain Pkinase and PK-Tyr-Ser-Thr domains, with FTA2 (TEY subgroup) and APH (TDY subgroup) domains restricted to specific members. Promoter analysis of 2-kb upstream regions (Figure 4E) identified 164 stress-responsive elements, including 5 defense-related elements in *SmMPK4/7/17* (with *SmMPK4* harboring an elicitor-responsive element). Hormone-responsive elements were abundant (78 total), particularly for MeJA (26) and ABA (29). Only 24 growth-related elements were detected, which were absent in 5 *SmMAPKs*. Gene structure analysis (Figure 4F) showed that TDY-subtype genes uniformly contain 10 exons, while TEY-subtype genes typically have 6 (except *SmMPK1/14* with 2). These results suggest that SmMAPKs primarily regulate stress and hormone responses, with structural divergence between subgroups.

A neighbor-joining phylogenetic tree of *Arabidopsis thaliana* (10) and *Salvia miltiorrhiza* (7) MAPKK proteins revealed close evolutionary relationships, with all SmMAPKKs nested within

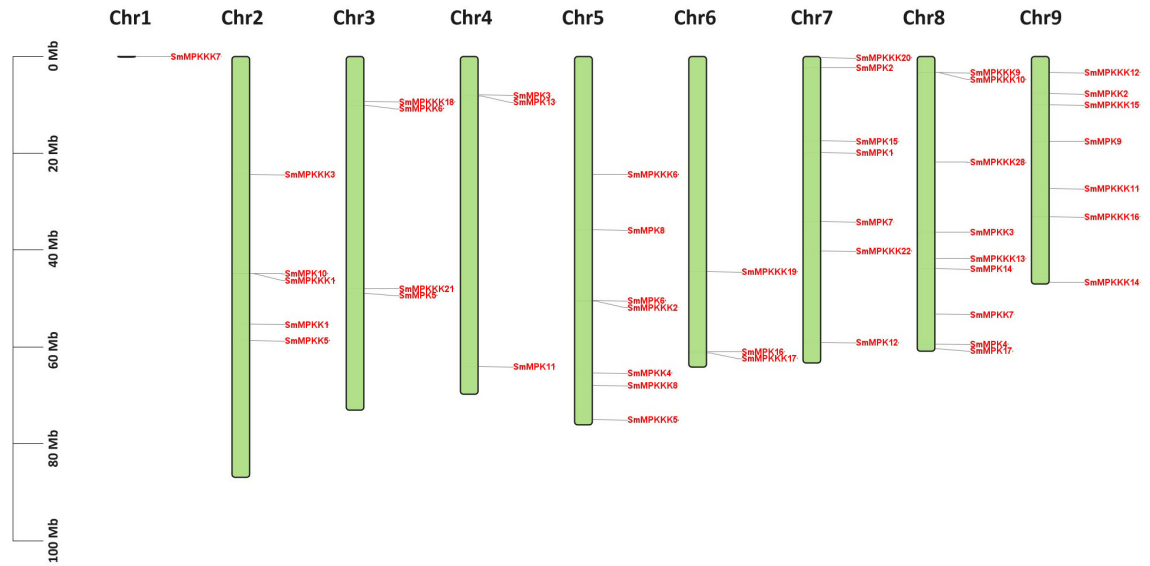


FIGURE 1
Chromosome mapping of the *S. miltiorrhiza* MAPK gene families. The scale on the left represents the chromosome lengths. The green boxes represent the chromosomes. MAPK gene members are marked in red.

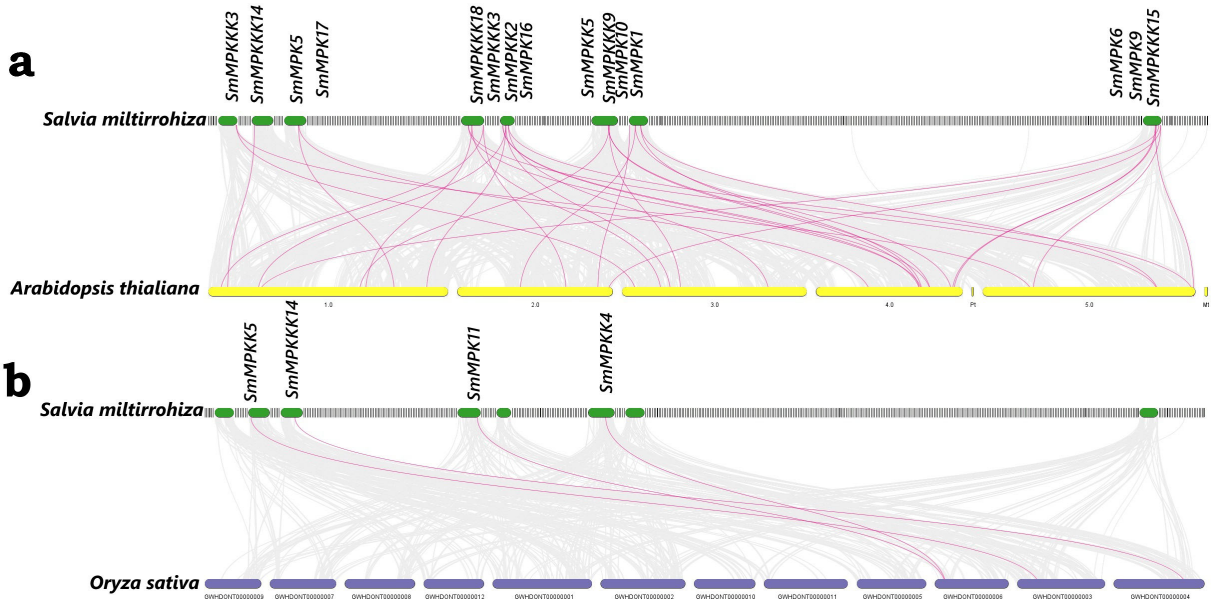


FIGURE 2
Synteny analysis of MAPK genes between **(A)** *S. miltiorrhiza* and *A. thaliana*; **(B)** *S. miltiorrhiza* and *O. sativa*. The red lines represent the syntenic MAPK gene pairs, while the gray lines represent the syntenic blocks within the genomes.

Arabidopsis clades (Figure 5A). Sequence alignments confirmed the conserved MAPKK motif [S/T]xxxxx[S/T] in all seven SmMAPKKs (Figure 5B). MEME analysis identified 10 motifs (Figure 5C; Supplementary Table S4), with subgroup-specific distributions: Group I (SmMAPKK1/2/6) contained N-terminal motif 7 but lacked motif 10, Groups II (SmMAPKK3) and III (SmMAPKK7) lacked both motifs, and Group IV (SmMAPKK4/5) exclusively contained motif 10. Domain analysis showed that all members harbor Pkinase and PK-Tyr-Ser-Thr domains, except SmMAPKK2, which uniquely possessed a DUF2764 domain (Figure 5D).

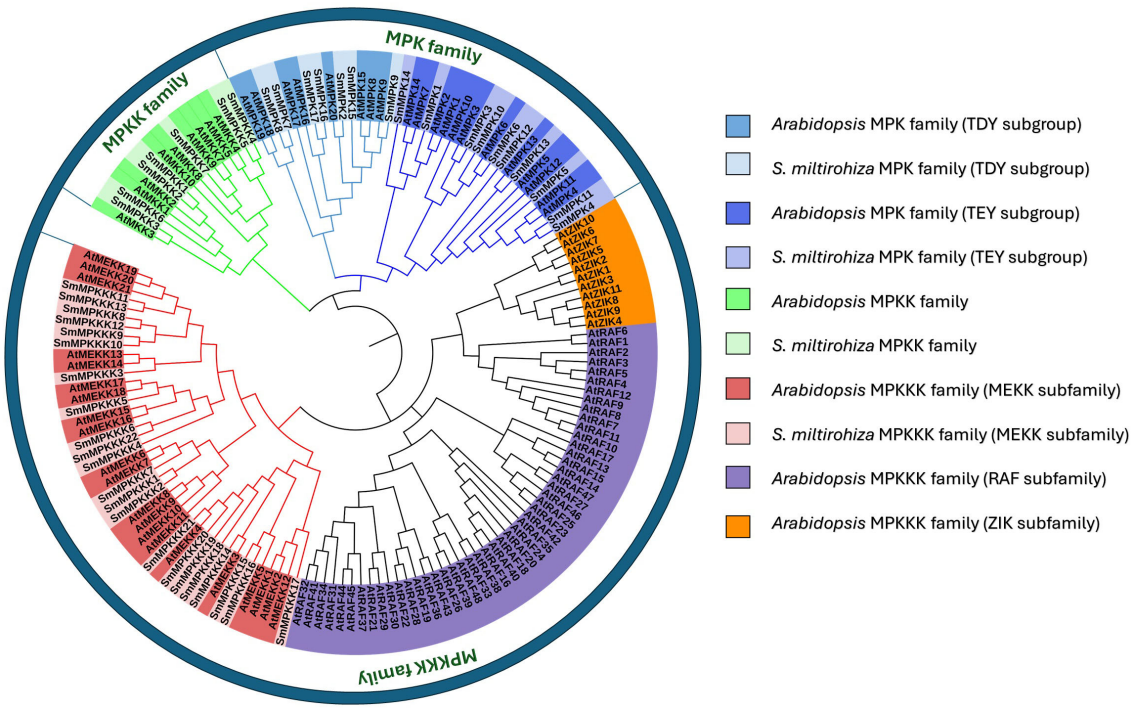


FIGURE 3
Phylogenetic tree of *S. miltiorrhiza* and *A. thaliana* MAPK gene families.

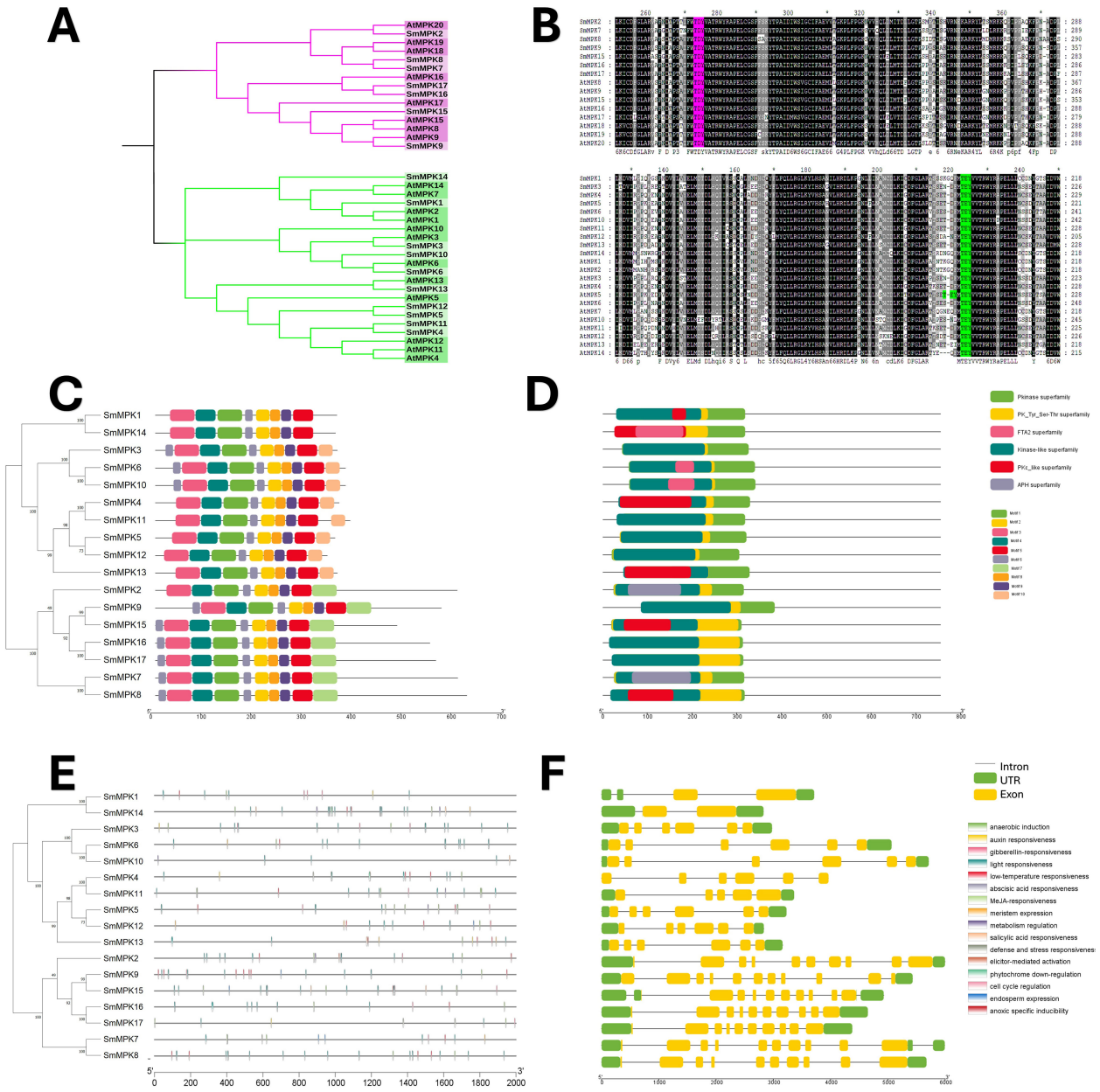


FIGURE 4
characterization analyses of *S. miltiorrhiza* MAPK gene family. **(A)** Phylogenetic tree of *S. miltiorrhiza* and *A. thaliana* MAPK proteins. The phylogenetic tree was done by the neighbor-joining (NJ) method with a 1,000-iteration bootstrap. TDY group with pink color and TEY group in green. **(B)** Alignments of *S. miltiorrhiza* and *A. thaliana* MAPK amino acid sequences. Conserved signatures are highlighted by different colors. **(C)** Conservative motif distribution of SmMAPK proteins. Boxes with different colors represent the 10 different motifs identified by meme suit. **(D)** Conserved domains of SmMAPK gene family. Boxes with different colors represent the different conserved domains identified. **(E)** Cis-elements in promoters of SmMAPK genes. Cis-acting elements with different functions are represented by different colored boxes. **(F)** The gene structure of SmMAPKs. Black lines represent introns, green boxes represent untranslated regions, and yellow boxes represent exons.

Promoter analysis detected 61 stress-responsive elements, including defense-related elements in *SmMAPKK2* (elicitor responsive) and *SmMAPKK5*, and 32 hormone-responsive elements (MeJA: 6; ABA: 12). Notably, *SmMAPKK2* lacked hormone-responsive elements. Gene structure analysis divided SmMAPKKs into two subgroups as follows: one with single-exon genes (*SmMAPKK4/5/7*) and another with multi-exon genes (*SmMAPKK1/2/3/6*) containing seven to eight exons (Figure 5F). These findings suggest functional diversification among SmMAPKK subgroups in stress and hormonal responses.

Phylogenetic analysis revealed close evolutionary relationships between *S. miltiorrhiza* MAPKKKs and *A. thaliana* MEKK subfamily (Figures 6A, B). The conserved MEKK motif G[T/S]Px [W/Y/F]MAPEV was present in all SmMAPKKKs except

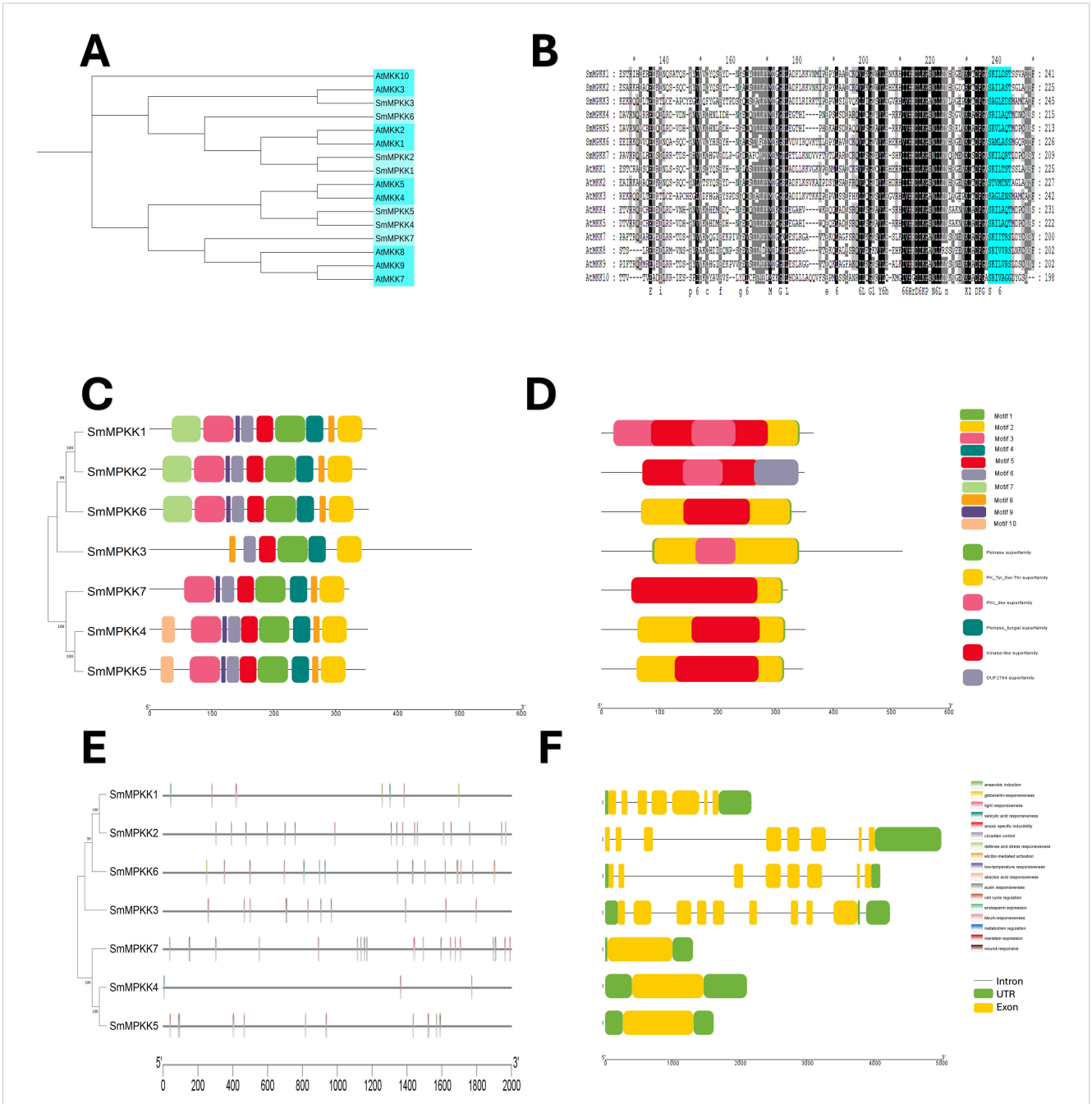


FIGURE 5
Characterization analyses of *S. miltiorrhiza* MAPKK gene family. **(A)** Phylogenetic analysis of *S. miltiorrhiza* and *A. thaliana* MAPKK proteins. The phylogenetic tree was done by the neighbor-joining (NJ) method with a 1,000-iteration bootstrap. **(B)** Alignments of amino acid sequences of *S. miltiorrhiza* and *A. thaliana* MAPKK gene family. Conserved signatures are highlighted by different colors. **(C)** Conservative motif distribution of SmMAPKK proteins. Boxes with different colors represent the 10 different motifs identified by meme suit. **(D)** Conserved domains of SmMAPKK gene family. Boxes with different colors represent the different conserved domains identified. **(E)** Cis-elements in promoters of SmMAPKK genes. Cis-acting elements with different functions are represented by different colored boxes. **(F)** The gene structure of SmMAPKKs. Black lines represent introns, green boxes represent untranslated regions, and yellow boxes indicate exons.

SmMAPKKK20. MEME analysis identified 10 motifs (Figure 6C; Supplementary Table S4), with subgroup-specific distributions as follows: Group I contained motif 9 (absent motif 4), while Groups II–IV retained motif 4 but lacked motif 9. All members harbored Pkinase and PK-Tyr-Ser-Thr domains (Figure 6D). Promoter analysis detected 261 stress-responsive elements (primarily light/low-temperature related), with 17 defense elements across 12 genes, including an elicitor-responsive element in *SmMPKKK11*. Hormone-responsive elements (196 total) were abundant, dominated by MeJA (92) and ABA (62) elements. Gene structure varied widely ranging from intronless (*SmMPKKK21/14/2/13/8*) to four-exon (*SmMPKKK10*) architectures (Figure 6F). These findings highlight SmMAPKKKs' roles in stress and hormonal signaling.

10.3389/fpls.2025.1583953

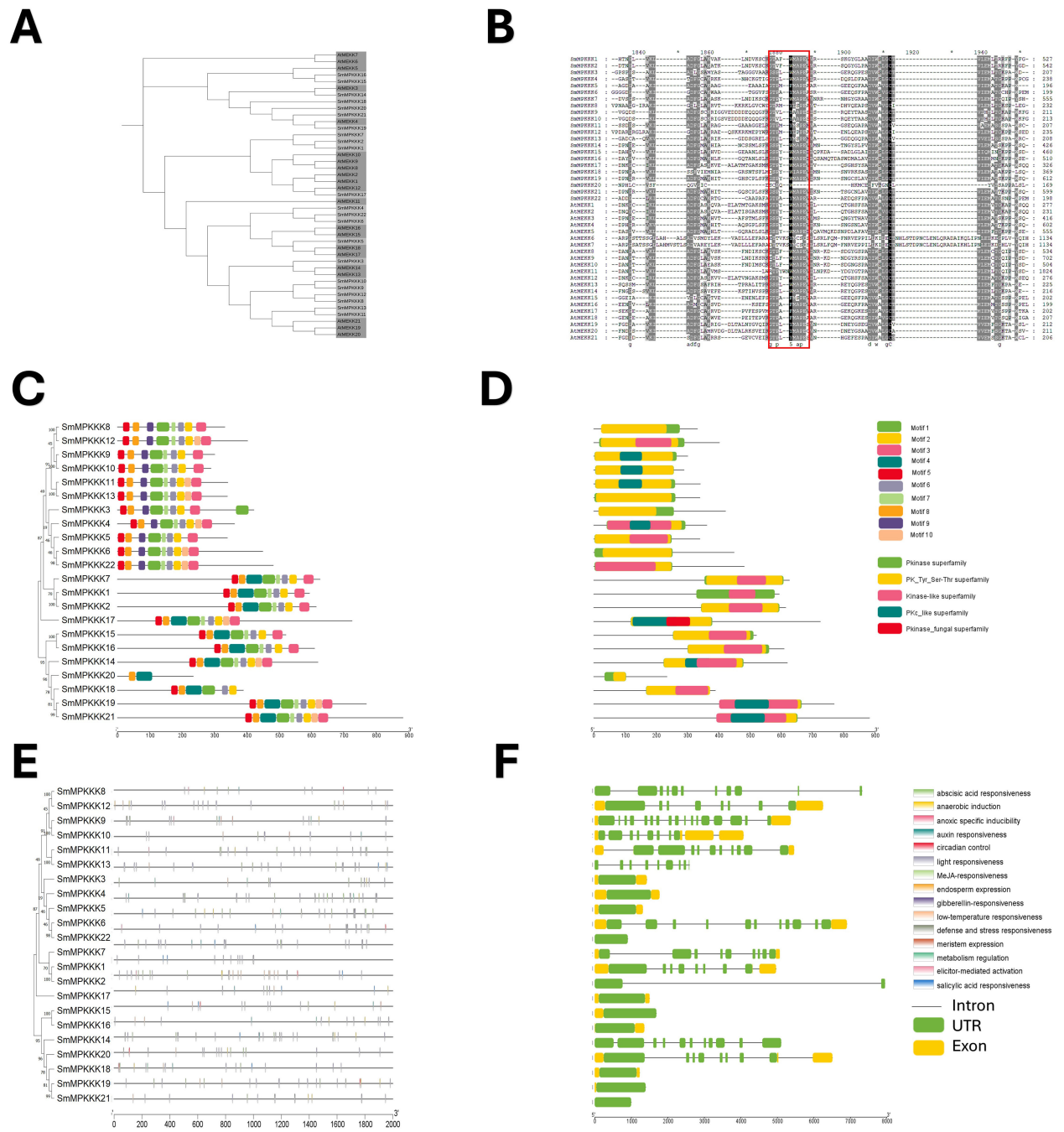


FIGURE 6
Characterization of *S. miltiorrhiza* MAPKKK gene family. **(A)** Phylogenetic tree of *S. miltiorrhiza* and *A. thaliana* MAPKKK proteins was generated using the neighbor-joining (NJ) method, and a bootstrap analysis was conducted with 1,000 iterations. **(B)** Alignments of amino acid sequences of *S. miltiorrhiza* and *A. thaliana* MAPKKK gene family. Conserved signatures are highlighted by different colors. **(C)** Conservative motif distribution of SmMAPKKK proteins. Boxes with different colors represent the 10 different motifs identified by meme suit. **(D)** Conserved domains of SmMAPKKK gene family. Boxes with different colors represent the different conserved domains identified. **(E)** Cis-elements in promoters of SmMAPKKK genes. Cis-acting elements with different functions are represented by different colored boxes. **(F)** The gene structure of SmMAPKKKs. Black lines represent introns, and green and yellow boxes represent untranslated regions and exons, respectively.

Protein-protein interaction network analysis

To gain insights into the functional relationships and potential signaling pathways involving the identified MAPK genes in *S. miltiorrhiza*, a protein–protein interaction (PPI) network was constructed using the *A. thaliana* MAPK orthologs as a reference. The PPI network (Figure 7) revealed significant interactions between *S. miltiorrhiza* MAPKs and their *Arabidopsis* counterparts highlighting conserved regulatory mechanisms across species.

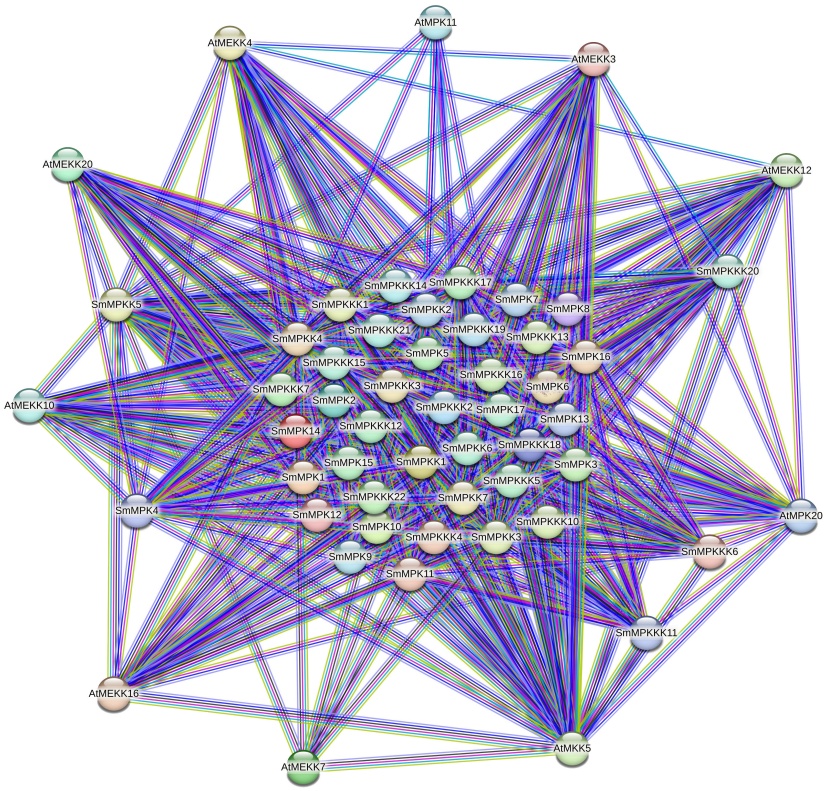


FIGURE 7
Protein–protein interaction (PPI) network illustrating interactions between the identified SmMAPKs proteins and their orthologs of *Arabidopsis*. Nodes represent proteins, and edges indicate known and predicted interactions. The thickness of the edges corresponds to the confidence score of the interaction. Different edge colors represent different interactions such as experimental data, co-expression, databases, or text mining.

Our results show that SmMPKK5 strongly interacts with AtMEKK10 (combined score: 0.854), an MAPKKK involved in ABA signaling and drought stress responses (Danquah et al., 2015; Xing et al., 2008). This suggests that SmMPKK5 may participate in ABA-mediated abiotic stress signaling in *S. miltiorrhiza*. Additionally, SmMPKKK6, SmMPKKK11, and SmMPKKK20 show interactions with AtMEKK12, AtMEKK16, and AtMEKK20, which are known regulators of PAMP-triggered immunity (PTI) and MAPK cascades responding to pathogen infection (Asai et al., 2002; Gao et al., 2008; Wang et al., 2017). For example, AtMEKK1 (closely related to AtMEKK12) initiates PTI signaling by activating downstream MAPKs like MPK3/6 in response to bacterial and fungal elicitors (Meng and Zhang, 2013). AtMEKK16 has also been associated with stress signal integration, including responses to wounding and oxidative stress (Ichimura et al., 2002a). These interactions suggest that the *S. miltiorrhiza* MAPKs may act within conserved stress signaling pathways regulating responses to both abiotic stresses (e.g., salt, drought) and biotic stimuli (e.g., fungal elicitors).

Gene Ontology enrichment analysis

To investigate the functional roles of MAPK cascade genes in *Salvia miltiorrhiza*, GO enrichment analysis was conducted revealing significant associations across cellular components, molecular function, and biological process categories (Figure 8). Most MAPK genes were localized to the cytoplasm (40 genes) and nucleus (23 genes), indicating their involvement in both signal transduction and gene regulation. In the biological process category, 35 genes were directly associated with the MAPK cascade, 43 with protein phosphorylation and regulation of cellular processes, and 39 with signal transduction, underscoring the widespread involvement of MAPK genes in signaling and cellular regulation. Additionally, eight genes were linked to defense responses highlighting the potential roles of these kinases in plant immunity and stress adaptation. Molecular function enrichment showed a strong bias toward kinase-related activities, with 43 genes exhibiting protein kinase activity and ATP binding, and 18 specifically annotated with MAP kinase activity. Notably, *SmMPK4* displayed enrichment in MAP kinase activity, ATP binding, and nucleotide interactions consistent with its role as a core MAPK in stress-responsive signaling. *SmMPKK5* was significantly associated with MAPK kinase activity and phosphotransferase functions positioning it as a key MAPKK that likely activates downstream MAPKs such as *SmMPK4*. *SmMPKKK6* exhibited strong enrichment in protein kinase and kinase-binding functions suggesting its role in initiating MAPK cascades by interacting with and activating MAPKKs. *SmMPKKK11* showed general kinase activity and ATP binding implying a more basal regulatory function, while

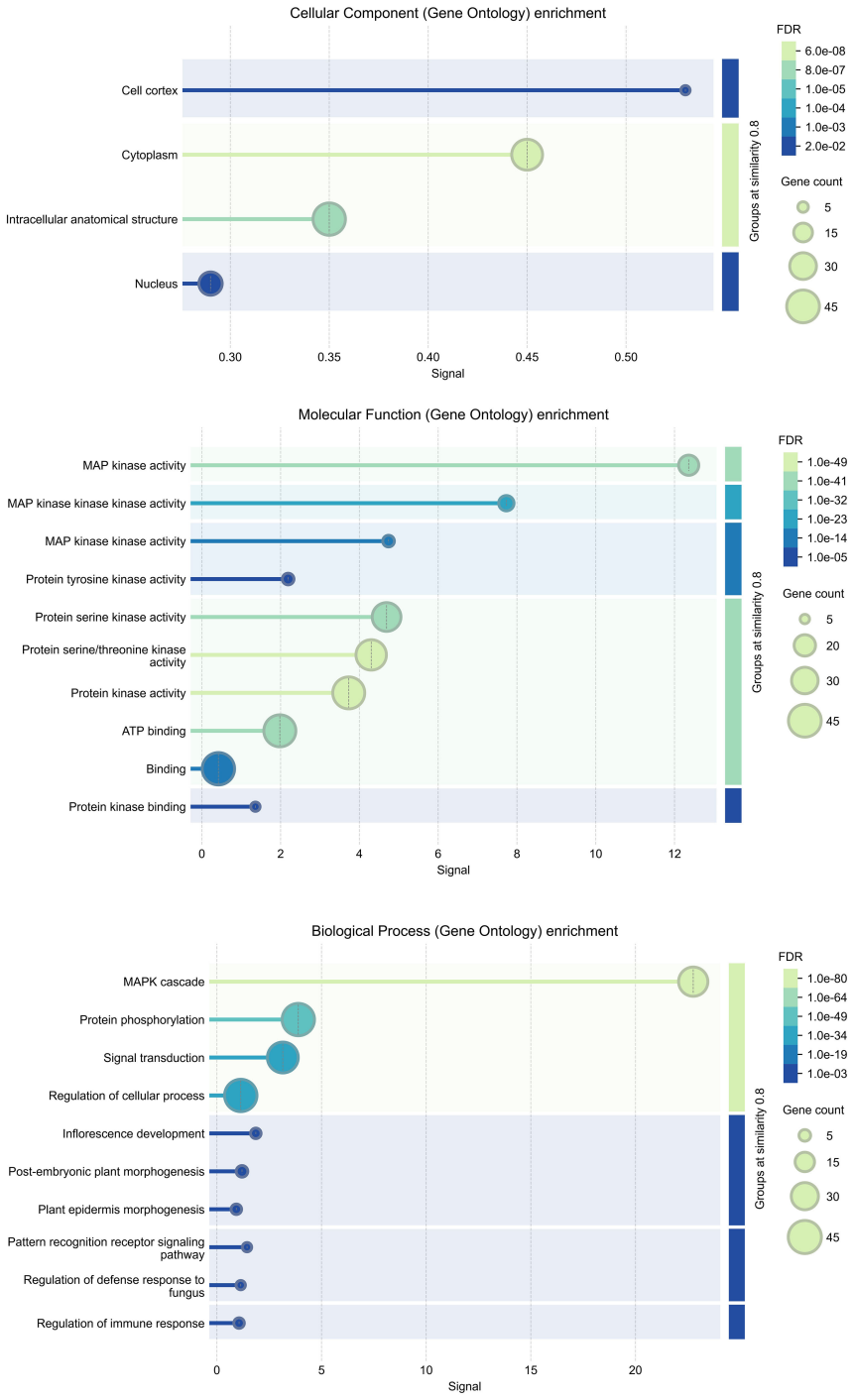


FIGURE 8
GO analysis of SmMAPKs genes. GO, Gene Ontology.

SmMPKKK20 was distinctly enriched in MAPKKK activity indicating a specialized role in transducing upstream signals from membrane-based stimuli like pathogen recognition or hormonal cues. All GO annotation processes are summarized in Supplementary Table S7.

KEGG pathway enrichment analysis

To further elucidate the functional roles of the identified MAPK genes in *Salvia miltiorrhiza*, KEGG pathway enrichment analysis was performed (Figure 9). The analysis revealed significant

enrichment in pathways related to signal transduction and plant–pathogen interactions providing insights into the biological processes regulated by these genes. The most significantly enriched pathway was the MAPK signaling pathway—plant, with 36 genes associated with this pathway. This finding confirms the central role of the identified MAPK genes in MAPK signaling, which is known to regulate various cellular processes, including stress responses, growth, and development. The enrichment of these genes in the MAPK signaling pathway underscores their importance in transmitting extracellular signals to intracellular responses, particularly in the context of fungal elicitor-mediated tanshinone accumulation. Another significantly enriched pathway was plant–pathogen interaction, with 13 genes associated with this pathway. This pathway is crucial for plant defense mechanisms against pathogens, and the enrichment of MAPK genes in this pathway suggests their involvement in plant immunity and stress responses. The presence of these genes in the plant–pathogen interaction pathway highlights their potential role in mediating defense responses to fungal elicitors, which may, in turn, regulate tanshinone biosynthesis.

Fungal elicitors promoted tanshinone accumulation in *S. miltiorrhiza*

To investigate the relationship between MAPK gene families and tanshinone induction by fungal elicitation, hairy roots were treated by two fungal elicitors: yeast extract and *A. niger* (Figure 10A). The results indicate that both elicitors do not have a significant effect on the fresh and dry weight of hairy roots (Figure 10B). However, yeast extract and *A. niger* significantly promoted the accumulation of dihydrotanshinone, cryptotanshinone, and miltirone in *S. miltiorrhiza* hairy roots, but no significant change in tanshinone I and tanshinone IA was observed. Moreover, the effect of yeast extract elicitor on the accumulation of tanshinones was higher than that of *A. niger* (Figure 10C).

Fungal elicitor-activated SmMAPK genes involved in tanshinone accumulation in *S. miltiorrhiza*

To identify fungal elicitor-activated SmMAPK genes that may be involved in tanshinone accumulation, hairy roots samples of yeast extract- and *A. niger*-treated groups, together with the control group, were collected before and after 1 and 4 days of treatments, then analyzed by RNA sequencing. All SmMAPK genes were screened in the transcriptome data, and their expression level was visualized using tbtool software to generate heatmaps (Figures 8, 11). Consistent with RNA-seq data, RT-qPCR (Figure 12) confirmed the significant upregulation of *SmMPK4* and *SmMPKK5*, while *SmMPKKK6*, *SmMPKKK11*, and *SmMPKKK20* were significantly suppressed in response to fungal elicitation. These results corroborate the reliability of our transcriptomic data and highlight key genes. To better understand the role of MAPK genes in fungal elicitor-mediated tanshinone accumulation, a Pearson correlation coefficient analysis was done. The results revealed that eight MPKs positively correlated with three enhanced tanshinone compounds (dihydrotanshinone, cryptotanshinone, and miltirone) with high negative correlation in *SmMPK9* and *SmMPK16* and high positive correlation of dihydrotanshinone, cryptotanshinone with *SmMPK4*, while 10 MPKs correlated negatively with them, especially *SmMPK1* and *SmMPK6*. Four of the seven identified SmMPKKs have a positive correlation with tanshinones, but only *SmMPKK3* and *SmMPKK5* showed a significant correlation. The other members showed medium-to-low negative correlation. Most SmMPKKK genes negatively correlated with tanshinones, and a very high correlation was observed in *SmMPKKK6*, *SmMPKKK11*, and *SmMPKKK20* (Figure 13).

Discussion

In recent decades, Chinese herbal medicines and phytometabolites have shown health-promoting actions and promising effects in enhancing immunity (Hao and Liu, 2022).

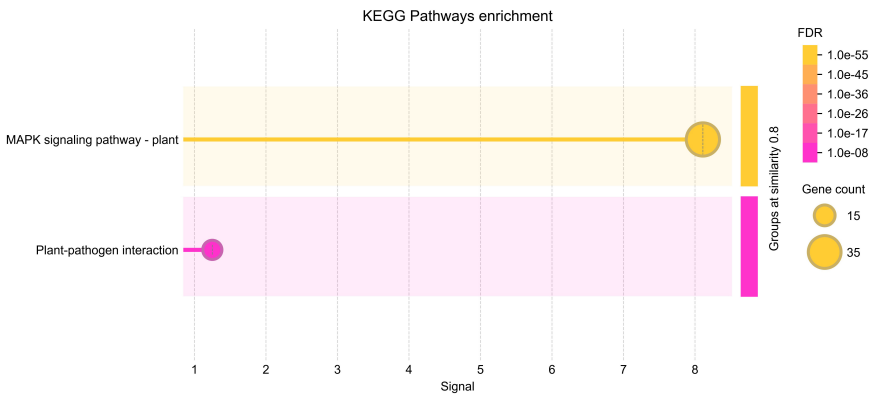


FIGURE 9
KEGG enrichment analysis of SmMAPKs genes. KEGG, Kyoto Encyclopedia of Genes and Genomes.

10.3389/fpls.2025.1583953

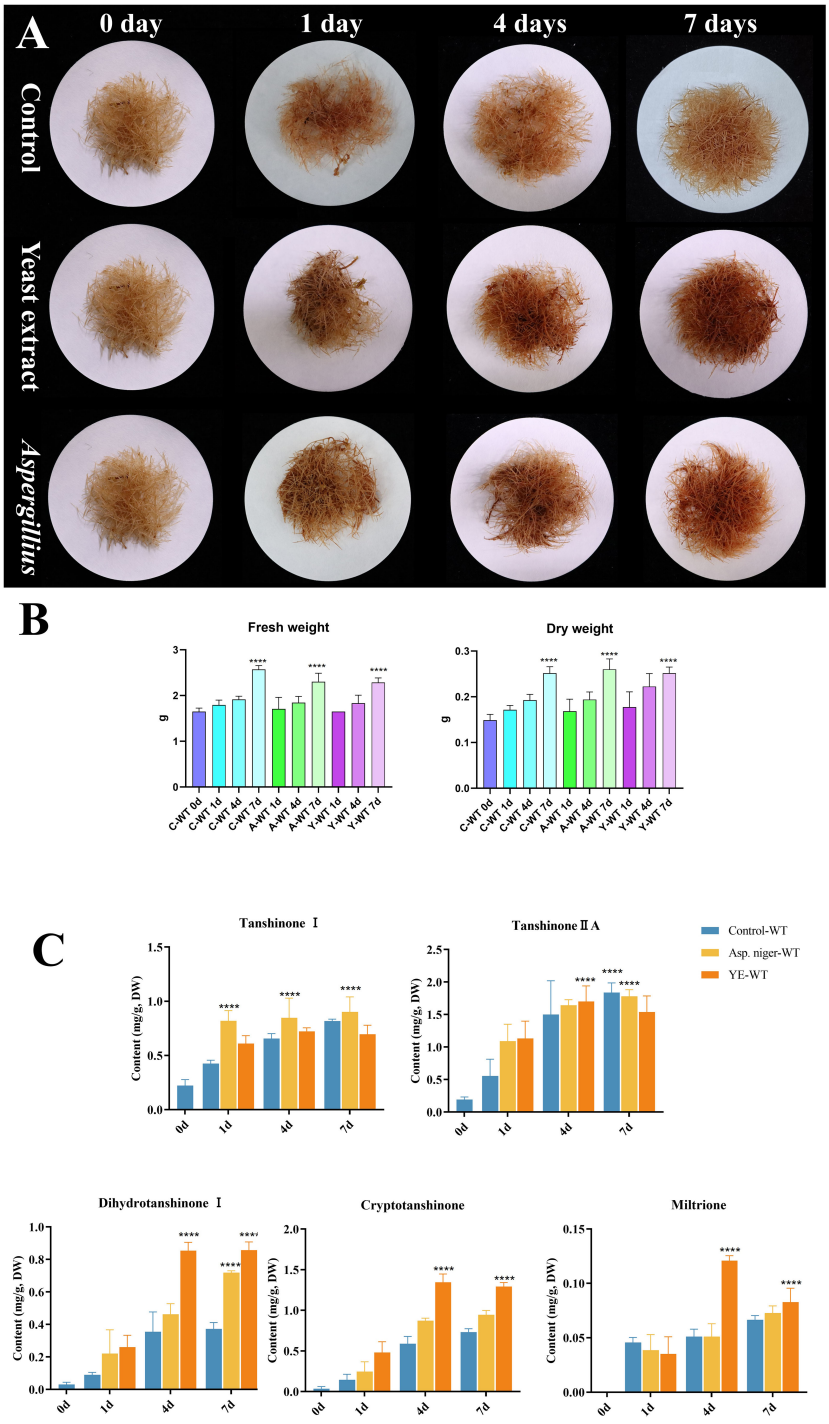


FIGURE 10
Effect of yeast extract and **(A)** *niger* elicitors on growth and tanshinone accumulation in *S. miltiorrhiza* hairy roots. **(A)** *S. miltiorrhiza* hairy roots treated by the two fungal elicitors for 1, 4, and 7 days. **(B)** Fresh and dry weight of *S. miltiorrhiza* hairy roots treated by the two fungal elicitors. **(C)** Tanshinone content in *S. miltiorrhiza* hairy roots treated by the two fungal elicitors. Error bars indicate the mean ± standard deviation (SD) based on three biological replicates. One-way ANOVA multiple comparison test was used to determine significant differences. (****) represents a significant difference at the level of $p < 0.0001$ relative to the control group (0-day treatment).

Genome-wide identification of gene families gives valuable information on the regulation process in plants (Wang et al., 2024). SmMAPK genome-wide identification, characterization, and phylogenetic analysis can provide useful information to enhance our understanding of their potential regulatory roles in tanshinone accumulation in *S. miltiorrhiza*. The *A. thaliana* genome contains 20 AtMAPKs, 10 AtMAPKKs, and 80 AtMAPKKKs, while 17 OsMAPKs, 8 OsMAPKKs, and 75

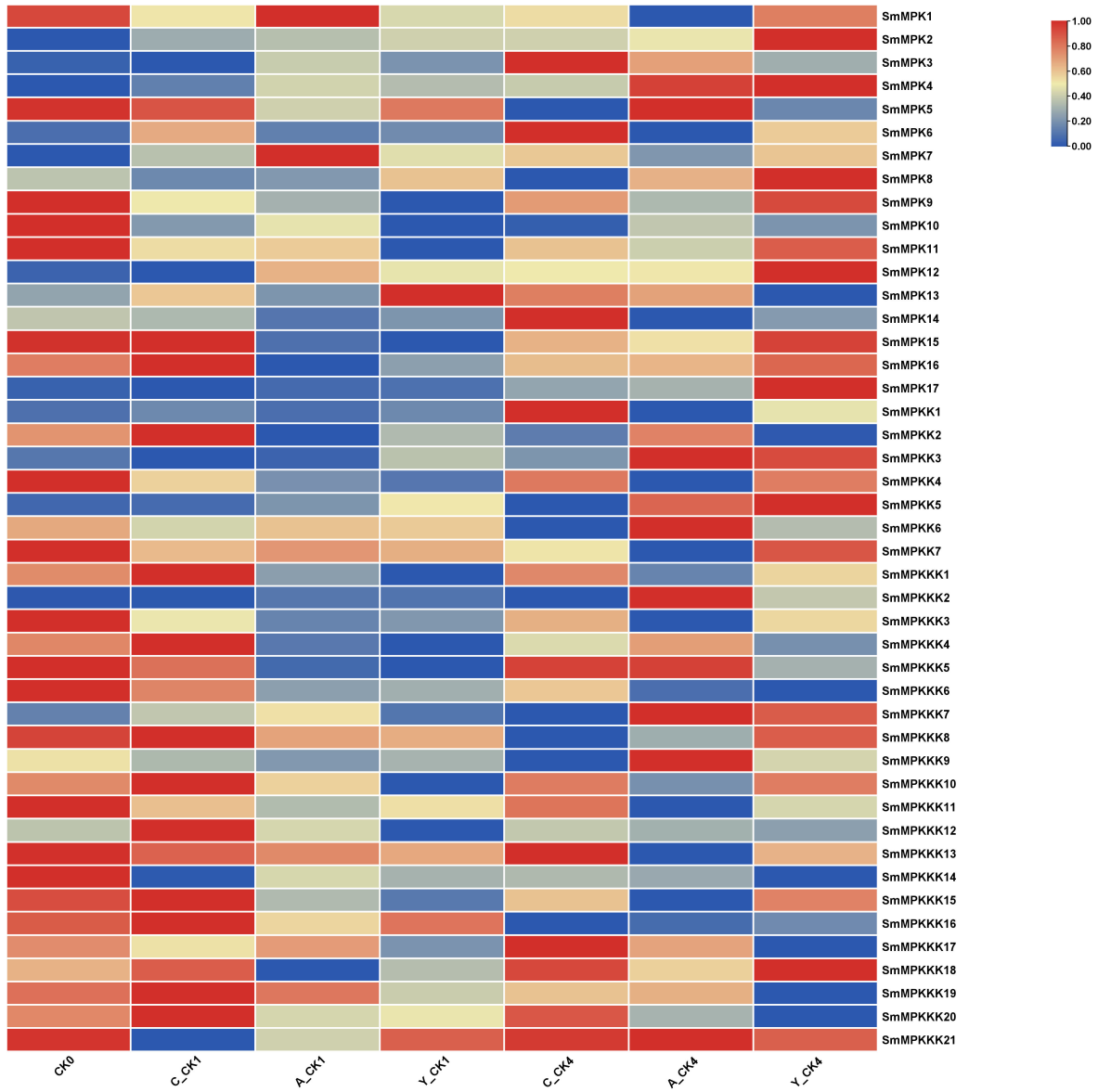


FIGURE 11
Heatmap of differentially expressed MAPK genes (DEGs) in fungal elicitor-treated hairy roots. Hierarchical clustering of normalized expression values (log2-transformed TPM/FPKM/variance-stabilized counts) for genes with significant differential expression ($|log2FC| \geq 0.5$, FDR-adjusted $p < 0.05$) between treatment and control. Rows represent genes; columns represent samples grouped by condition. Expression levels are scaled by row and depicted as a color gradient (blue: downregulated; red: upregulated; yellow: mean expression). C_CK0, control group before treatment; C_CK1, control group 1 day after treatment; C_CK4, control group 4 days after treatment; A_CK1, *Aspergillus*-treated hairy roots for 1 day; A_CK4, *Aspergillus*-treated hairy roots for 4 days; Y_CK1, yeast-treated hairy roots for 1 day; Y_CK4, yeast-treated hairy roots for 4 days.

OsMAPKKKs were detected in the rice genome (Rao et al., 2010). In this study, 17 MAPKs, 7 MAPKKs, and 22 MAPKKKs were identified in *S. miltiorrhiza* genome, which were distributed on nine chromosomes, with MWs ranging from 25,860.58 to 95,481.2 Da and theoretical pI ranging from 4.51 to 9.52.

MAPK genes are important in various cellular processes, including development, growth, cell death, and abiotic and biotic stress response (Smékalová et al., 2014; Guan et al., 2014; Wang et al., 2018; Zhang et al., 2017; Cai et al., 2014). A very significant number of stress-responsive and phytohormone elements were detected in the promoter sites of SmMAPK genes, in addition to some development and growth cis-acting elements. That indicates that SmMAPK genes are mostly related to biotic and abiotic stress as well as phytohormone response. Moreover, the identification of 5 defense-response elements in *SmMPK4*, *SmMPK7*, and *SmMPK17* of which *SmMPK4* has an elicitor-mediated activation element (Figure 4E), 2 defense elements in *SmMPKK2*, which has 1 elicitor-mediated activation element and 1 defense element in *SmMPKK5* (Figure 5E), in addition to 17 defense-response elements, were distributed within *SmMPKKK6*, *SmMPKKK2*, *SmMPKKK14*, *SmMPKKK8*, *SmMPKKK9*, *SmMPKKK11*, *SmMPKKK5*, *SmMPKKK15*, *SmMPKKK20*, *SmMPKKK19*,

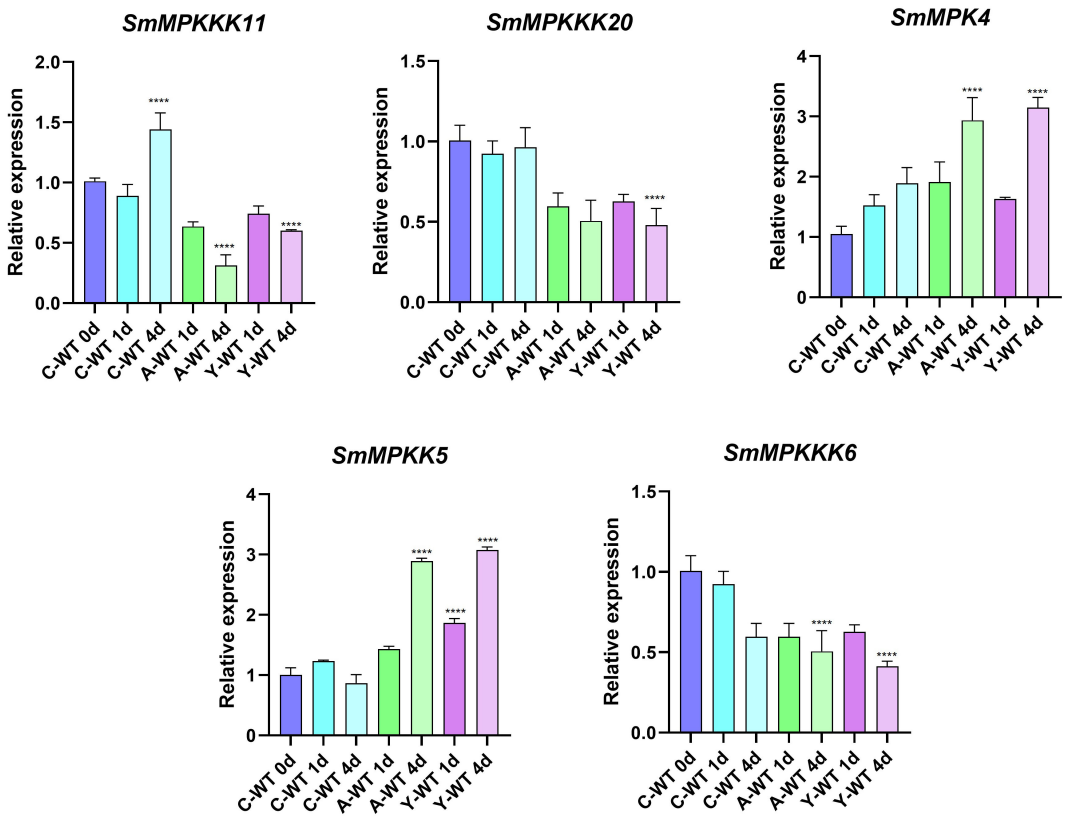


FIGURE 12
qRT-PCR validation of expression level of candidate SmMAPK genes. Error bars represent the mean ± standard deviation (SD) of three biological replicates. Significant differences were detected with unpaired *t* test. (****) represents a significant difference at the level of $p < 0.0001$ compared with the control group.

SmMPKKK21, and *SmMPKKK22*, of which *SmMPKKK11* also contains one elicitor-mediated activation element (Figure 6E), which suggest these members to be possible candidate genes involved in fungal elicitor-mediated tanshinone accumulation.

The divergence in exon/intron structure is crucial to the evolution of gene families. Thus, the structure of gene exons and introns can support phylogenetic groupings (Shiu and Bleecker, 2003; Zhang et al., 2012; Cao et al., 2011). The structures of the SmMAPK genes were predicted from the *S. miltirrohiza* genome. All SmMAPKs genes from the TDY subgroup have 10 exons, while most TEY subgroup genes have six exons (Figure 4F). This high similarity degree was observed among subgroup members and supports their relationships. Similarly, SmMAPKK genes were divided into the following two subgroups: the first subgroup with one exon and the second subgroup with seven to eight exons (Figure 5F). In contrast, the number of exons in SmMPKKK genes subgroups varies from one to four exons, while some members have no exons (Figure 6F). This large variation in structures may be due to significant changes in the genome during the long evolutionary history.

The protein interaction network analysis (Figure 7) revealed that MAPK genes are involved in complex signaling pathways that are likely to play a role in the accumulation of tanshinones in response to fungal elicitors. The interactions with *Arabidopsis* orthologs, particularly those involved in stress and defense responses, suggest that these genes may be key regulators in the signaling pathways that lead to the production of tanshinones.

GO enrichment analysis (Figure 8) reveals that the MAPK genes in *S. miltiorrhiza* are involved in a wide range of cellular processes, particularly those related to stress responses, signal transduction, and secondary metabolite biosynthesis. These findings provide valuable insights into the functional roles of these genes and their potential applications in improving tanshinone production.

The KEGG pathway enrichment analysis (Figure 9) provides valuable insights into the functional roles of the MAPK genes identified in *Salvia miltiorrhiza*. The results highlight the involvement of these genes in critical signaling pathways, particularly those related to stress responses and plant–pathogen interactions. The enrichment of MAPK genes in both the MAPK signaling pathway and plant–pathogen interaction pathway suggests that these genes are not only involved in stress responses but also play a role in regulating secondary metabolite biosynthesis. The activation of these genes in response to fungal elicitors may lead to the upregulation of biosynthetic pathways involved in tanshinone production. This provides a foundation for further studies aimed at manipulating these genes to enhance tanshinone production through biotechnological approaches.

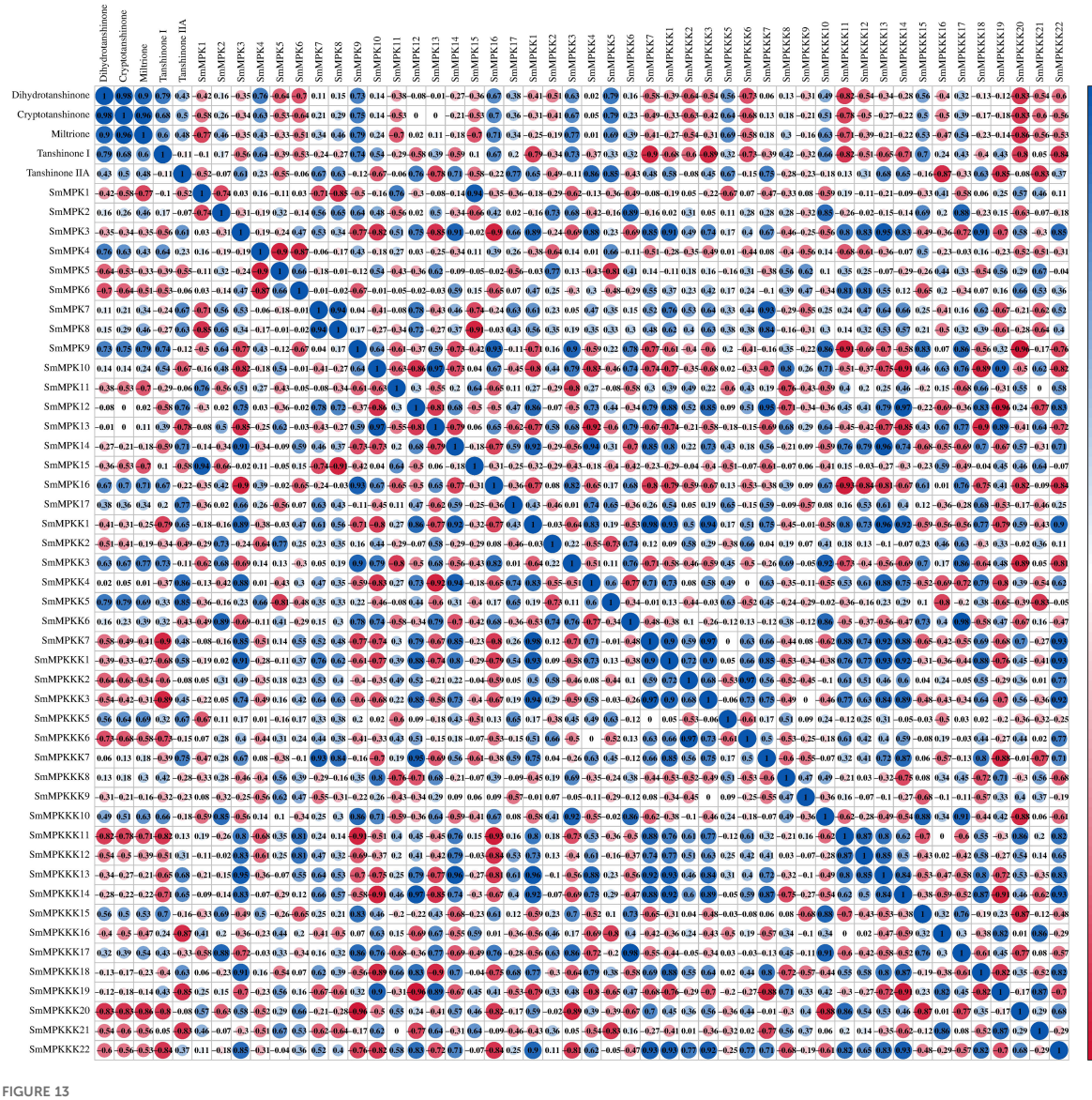


FIGURE 13
Pearson correlation coefficient analysis to screen the potential SmMAPK genes involved in tanshinone accumulation in *S. miltiorrhiza*. The circle size reflects the strength of the association between the gene expression level and tanshinone content; red and blue circles mean a negative and a positive correlation, respectively. No circles means no correlation.

Tanshinone accumulation in *S. miltiorrhiza* was reported to be significantly induced by fungal elicitors (Zhou et al., 2017; Wu et al., 2022) suggesting that they are a part of defensive mechanisms, especially that tanshinone compounds were reported to have antimicrobial properties. Moreover, cryptotanshinone and dihydrotanshinone I showed higher antimicrobial activity than tanshinone IIA and tanshinone I (Zhao et al., 2011). In our results (Figure 10C), both fungal elicitors yeast extract and *A. niger* increased specific tanshinone compound accumulation (miltirone, cryptotanshinone, and dihydrotanshinone I), while no significant increase in tanshinone IIA and tanshinone I was observed. The increase in tanshinone compounds with higher antimicrobial properties further supports that their accumulation was induced as a defensive mechanism. Many studies showed that fungal-responsive MAPK genes have a regulatory role in defensive metabolite accumulation in plants (Kishi-Kaboshi et al., 2010). For example, in *Arabidopsis*, camalexin, the major phytoalexin, is regulated by two fungal-responsive MPK3/MPK6 (Ren et al., 2008). In rice, OsMPK3 and OsMPK6, the orthologs of *Arabidopsis AtMPK3* and AtMPK6, regulate the biosynthesis of momilactones and phytocassanes, the diterpenoid phytoalexins (Kishi-Kaboshi et al., 2010). Interactions between *MKK1*, *MKK4*, *MKK5*, *MAPK3*, and *MAPK6* with *MKK9* in *Brassica rapa* play a major role in camalexin biosynthesis in response to *Alternaria brassicae* (Gaur et al., 2018). The application of yeast extract dramatically increased biphenyl phytoalexins, the defensive metabolites in *Sorbus aucuparia*, by activating the MAPK pathway (Li et al., 2024). Our results (Figure 11) revealed upregulation or downregulation of most SmMAPK genes in response to the two fungal elicitors. Moreover, in the Pearson correlation coefficient analysis (Figure 13), *SmMPK4* showed a high positive correlation

with dihydrotanshinone and cryptotanshinone, the two tanshinones with higher antimicrobial properties. In addition, the promoter analysis of *SmMPK4* (Figure 4E) identified two defense-response elements and one elicitor-mediated activation element. Also, the phylogenetic tree (Figure 3) showed that *SmMPK4* is the closest ortholog to *Arabidopsis AtMPK4*, a kinase protein involved in defense response and regulation of phytoalexin accumulation (Brodersen et al., 2006; Droillard et al., 2004; Petersen et al., 2000; Schweighofer et al., 2007; Qiu et al., 2008). Overall, these findings indicate *SmMPK4* to be a suitable candidate that has a regulatory role in tanshinone accumulation. *SmMPKK5* has a high positive correlation with both tanshinone compounds and *SmMPK4*. In addition, it has a defense element in the promoter site suggesting that it also may have a regulatory role. *AtMPKK2* is a MAPK kinase required for the activation of *AtMPK4* (Kong et al., 2012). Our results showed that *SmMPKK2*, which is the closest ortholog of *AtMPKK2*, significantly correlated with *SmMPK4*, and its promoter site has two defense elements and one elicitor-mediated activation element. This may suggest its possible regulatory role in tanshinone accumulation. *SmMPKKK6*, *SmMPKKK11*, and *SmMPKKK20* have a very high negative correlation with tanshinone compounds. In addition, the three member promoter sites have defense cis-regulatory elements, and *SmMPKKK11* has one elicitor-mediated activation element suggesting their involvement in fungal elicitor-mediated tanshinone accumulation in *S. miltiorrhiza*.

In conclusion, our findings on the potential roles of MAPK in tanshinone accumulation are consistent with previous studies in *Arabidopsis* and rice, where MAPKs have been shown to regulate the biosynthesis of defensive metabolites such as camalexin and momilactones (Kishi-Kaboshi et al., 2010; Gaur et al., 2018). These similarities suggest a conserved regulatory mechanism across plant species. The identification of key MAPK genes involved in tanshinone biosynthesis opens new avenues for biotechnological applications, such as the use of genetic engineering or fungal elicitors to enhance tanshinone production in *S. miltiorrhiza*.

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://bigd.big.ac.cn/gsa/browse/CRA016503, CRA016503.

Author contributions

AA: Methodology, Software, Writing – original draft, Writing – review & editing. XD: Formal Analysis, Writing – review & editing. LaZ: Methodology, Writing – review & editing. FY: Investigation, Writing – review & editing. JW: Software, Validation, Writing – review & editing. LiZ: Funding acquisition, Resources, Supervision, Writing – review & editing. QC: Resources, Software, Visualization, Writing – review & editing. ZY: Project administration, Resources, Supervision, Writing – review & editing. DY: Formal Analysis, Funding acquisition, Project administration, Supervision, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This research was funded by the Key Scientific and Technological Grant of Zhejiang for Breeding New Agricultural Varieties (No. 2021C02074), the Key Project of the Central Government: Capacity Building for the Sustainable Utilization of Traditional Chinese Medicine Resources (No. 2060302), and the Scientific Research Start-up Fund of Zhejiang Sci-Tech University (XJ2023001002).

Acknowledgments

The authors gratefully acknowledge the financial support from the Key Scientific and Technological Grant of Zhejiang for Breeding New Agricultural Varieties and the Key Project of the Central Government: Capacity Building for the Sustainable Utilization of Traditional Chinese Medicine Resources.

Conflict of interest

Author DY was employed by the company Shaoxing Biomedical Research Institute of Zhejiang Sci-Tech University Co., Ltd.

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

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.2025.1583953/full#supplementary-material.

References

Arthur, J. S., and Ley, S. C. (2013). Mitogen-activated protein kinases in innate immunity. *Nat. Rev. Immunol.* 13, 679–692. doi: 10.1038/nri3495

Asai, T., Tena, G., Plotnikova, J., Willmann, M. R., Chiu, W. L., Gomez-Gomez, L., et al. (2002). MAP kinase signalling cascade in *Arabidopsis* innate immunity. *Nature* 415, 977–983. doi: 10.1038/415977a

Berardini, T. Z., Reiser, L., Li, D., Mezheritsky, Y., Muller, R., Strait, E., et al. (2015). The Arabidopsis information resource: Making and mining the "gold standard" annotated reference plant genome. *Genesis* 53, 474–485. doi: 10.1002/dvg.22877

Brodersen, P., Petersen, M., Bjørn Nielsen, H., Zhu, S., Newman, M. A., Shokat, K. M., et al. (2006). Arabidopsis MAP kinase 4 regulates salicylic acid- and jasmonic acid/ethylene-dependent responses via EDS1 and PAD4. *Plant J.* 47, 532–546. doi: 10.1111/j.1365-313X.2006.02806.x

Cai, G., Wang, G., Wang, L., Liu, Y., Pan, J., and Li, D. (2014). A maize mitogen-activated protein kinase kinase, ZmMKK1, positively regulated the salt and drought tolerance in transgenic Arabidopsis. *J. Plant Physiol.* 171, 1003–1016. doi: 10.1016/j.jplph.2014.02.012

Cao, J., Huang, J., Yang, Y., and Hu, X. (2011). Analyses of the oligopeptide transporter gene family in poplar and grape. *BMC Genomics* 12, 465. doi: 10.1186/1471-2164-12-465

Chen, C. J., Wu, Y., Li, J. W., Wang, X., Zeng, Z. H., Xu, J., et al. (2023). TBtools-II: A one for all, all for one bioinformatics platform for biological big-data mining. *Mol. Plant* 16, 1733–1742. doi: 10.1016/j.molp.2023.09.010

Contreras, A., Leroy, B., Mariage, P. A., and Wattiez, R. (2019). Proteomic analysis reveals novel insights into tanshinones biosynthesis in Salvia miltiorrhiza hairy roots. *Sci. Rep.* 9, 5768. doi: 10.1038/s41598-019-42164-3

Danquah, A., de Zelicourt, A., Colcombet, J., and Hirt, H. (2015). The role of ABA and MAPK signaling pathways in plant abiotic stress responses. *Biotechnol. Adv.* 32, 40–52. doi: 10.1016/j.bioteChadv.2013.09.006

Droillard, M. J., Boudsocq, M., Barbier-Brygoo, H., and Laurière, C. (2004). Involvement of MPK4 in osmotic stress response pathways in cell suspensions and plantlets of Arabidopsis thaliana: activation by hypoosmolarity and negative role in hyperosmolarity tolerance. *FEBS Lett.* 574, 42–48. doi: 10.1016/j.febslet.2004.08.001

Gao, M., Liu, J., Bi, D., Zhang, Z., Cheng, F., Chen, S., et al. (2008). MEKK1, MKK1/MKK2 and MPK4 function together in a mitogen-activated protein kinase cascade to regulate innate immunity in plants. *Cell Res.* 18, 1190–1198. doi: 10.1038/cr.2008.300

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., et al. (2005). "Protein identification and analysis tools on the ExPASy server," in *The proteomics protocols handbook*. Ed. J. M. Walker (Humana, Totowa, NJ), 571–607.

Gaur, M., Tiwari, A., Chauhan, R. P., Pandey, D., and Kumar, A. (2018). Molecular modeling, docking and protein-protein interaction analysis of MAPK signalling cascade involved in Camalexin biosynthesis in Brassica rapa. *Bioinformation* 14, 145–152. doi: 10.6026/97320630014145

Guan, Y., Lu, J., Xu, J., McClure, B., and Zhang, S. (2014). Two mitogen-activated protein kinases, MPK3 and MPK6, are required for funicular guidance of pollen tubes in Arabidopsis. *Plant Physiol.* 165, 528–533. doi: 10.1104/pp.113.231274

Guo, Z., Kang, S., Wu, Q., Wang, S., Crickmore, N., and Zhou, X. (2021). The regulation landscape of MAPK signaling cascade for thwarting Bacillus thuringiensis infection in an insect host. *PloS Pathog.* 17, e1009917. doi: 10.1371/journal.ppat.1009917

Hao, D., and Liu, C. (2022). Deepening insights into food and medicine continuum within the context of pharmacophylogeny. *Chin. Herbal Med.* 15, 1–2. doi: 10.1016/j.chmed.2022.12.001

Huang, J., Zhang, J., Sun, C., Yang, R., Sheng, M., Hu, J., et al. (2024). Adjuvant role of Salvia miltiorrhiza bunge in cancer chemotherapy: A review of its bioactive components, health-promotion effect and mechanisms. *J. Ethnopharmacology* 318, 117022. doi: 10.1016/j.jep.2023.117022

Ibraheem, F., Gaffoor, I., Tan, Q., Shyu, C. R., and Chopra, S. (2015). A sorghum MYB transcription factor induces 3-deoxyanthocyanidins and enhances resistance against leaf blights in maize. *Molecules* 20, 2388–2404. doi: 10.3390/molecules20022388

Ichimura, K., Casais, C., Peck, S. C., and Shinozaki, K. (2002a). MEKK1 is required for MPK4 activation and regulates tissue-specific and temperature-dependent cell death in *Arabidopsis*. *J. Biol. Chem.* 277, 329–335. doi: 10.1074/jbc.M103390200

Ichimura, K., Shinozaki, K., Tena, G., Sheen, J., Henry, Y., Champion, A., et al. (2002b). Mitogen-activated protein kinase cascades in plants: a new nomenclature. *Trends Plant Sci.* 7, 301–308. doi: 10.1016/S1360-1385(02)02302-6

Kawahara, Y., de la Bastide, M., Hamilton, J. P., Kanamori, H., McCombie, W. R., Ouyang, S., et al. (2013). Improvement of the Oryza sativa Nipponbare reference genome using next generation sequence and optical map data. *Rice* 6, 4. doi: 10.1186/1939-8433-6-4

Khan, A., Shah, S. T., Basit, A., Mohamed, H. I., and Li, Y. (2024). Mitogen-activated protein kinase: a potent signaling protein that combats biotic and abiotic stress in plants. *J. Plant Growth Regul.* 43, 1762–1786. doi: 10.1007/s00344-024-11239-5

Kishi-Kaboshi, M., Okada, K., Kurimoto, L., Murakami, S., Umezawa, T., Shibuya, N., et al. (2010a). A rice fungal MAMP-responsive MAPK cascade regulates metabolic flow to antimicrobial metabolite synthesis. *Plant J.* 63, 599–612. doi: 10.1111/j.1365-313X.2010.04264.x

Kishi-Kaboshi, M., Takahashi, A., and Hirochika, H. (2010b). MAMP-responsive MAPK cascades regulate phytoalexin biosynthesis. *Plant Signaling Behav.* 5, 1653–1656. doi: 10.4161/psb.5.12.13982

Kong, Q., Qu, N., Gao, M., Zhang, Z., Ding, X., Yang, F., et al. (2012). The MEKK1-MKK1/MKK2-MPK4 kinase cascade negatively regulates immunity mediated by a mitogen-activated protein kinase kinase kinase in Arabidopsis. *Plant Cell* 24, 2225–2236. doi: 10.1105/tpc.112.097253

Lee, S. R., Jeon, H., Kwon, J. E., Suh, H., Kim, B. H., Yun, M. K., et al. (2020). Anti-osteoporotic effects of Salvia miltiorrhiza Bunge EtOH extract both in ovariectomized and naturally menopausal mouse models. *J. Ethnopharmacology* 258, 112874. doi: 10.1016/j.jep.2020.112874

Li, Y., Yang, J., Zhou, J., Wan, X., Liu, J., Wang, S., et al. (2024). Multi-omics revealed molecular mechanism of biphenyl phytoalexin formation in response to yeast extract-induced oxidative stress in Sorbus aucuparia suspension cells. *Plant Cell Rep.* 43, 62. doi: 10.1007/s00299-024-03155-5

Meng, X., and Zhang, S. (2013). MAPK cascades in plant disease resistance signaling. *Annu. Rev. Phytopathol.* 51, 245–266. doi: 10.1146/annurev-phyto-082712-102314

Ming, Q., Su, C., Zheng, C., Jia, M., Zhang, Q., Zhang, H., et al. (2013). Elicitors from the endophytic fungus Trichoderma atroviride promote Salvia miltiorrhiza hairy root growth and tanshinone biosynthesis. *J. Exp. Bot.* 64, 5687–5694. doi: 10.1093/jxb/ert342

Pan, X., Chang, Y., Li, C., Qiu, X., Cui, X., Meng, F., et al. (2023). Chromosome-level genome assembly of Salvia miltiorrhiza with orange roots uncovers the role of Sm2OGD3 in catalyzing 15,16-dehydrogenation of tanshinones. *Horticulture Res.* 10, uhad069. doi: 10.1093/hr/uhad069

Petersen, M., Brodersen, P., Naested, H., Andreasson, E., Lindhart, U., Johansen, B., et al. (2000). Arabidopsis MAP kinase 4 negatively regulates systemic acquired resistance. *Cell* 103, 1111–1120. doi: 10.1016/S0092-8674(00)00213-0

Qiu, J. L., Fiil, B. K., Petersen, K., Nielsen, H. B., Botanga, C. J., Thorgrimsen, S., et al. (2008). Arabidopsis MAP kinase 4 regulates gene expression through transcription factor release in the nucleus. *EMBO J.* 27, 2214–2220. doi: 10.1038/emboj.2008.147

Rao, K. P., Richa, T., Kumar, K., Raghuram, B., and Sinha, A. K. (2010). In silico analysis reveals 75 members of mitogen-activated protein kinase kinase kinase gene family in rice. *DNA Res.* 17 (3), 139–153. doi: 10.1093/dnares/dsq011

Ren, J., Fu, L., Nile, S. H., Zhang, J., and Kai, G. (2019). Salvia miltiorrhiza in treating cardiovascular diseases: A review on its pharmacological and clinical applications. *Front. Pharmacol.* 10, 753. doi: 10.3389/fphar.2019.00753

Ren, D., Liu, Y., Yang, K. Y., Han, L., Mao, G., Glazebrook, J., et al. (2008). A fungal-responsive MAPK cascade regulates phytoalexin biosynthesis in Arabidopsis. *Proc. Natl. Acad. Sci. U.S.A.* 105, 5638–5643. doi: 10.1073/pnas.0711301105

Roux, P. P., and Blenis, J. (2004). ERK and p38 MAPK-activated protein kinases: a family of protein kinases with diverse biological functions. *Microbiol. Mol. Biol. Rev.* 68, 320–344. doi: 10.1128/MMBR.68.2.320-344.2004

Savojardo, C., Martelli, P. L., Fariselli, P., Profiti, G., and Casadio, R. (2018). BUSCA: an integrative web server to predict subcellular localization of proteins. *Nucleic Acids Res.* 46, W459–W466. doi: 10.1093/nar/gky320

Schweighofer, A., Kazanaviciute, V., Scheikl, E., Teige, M., Doczi, R., Hirt, H., et al. (2007). The PP2C-type phosphatase AP2C1, which negatively regulates MPK4 and MPK6, modulates innate immunity, jasmonic acid, and ethylene levels in Arabidopsis. *Plant Cell* 19, 2213–2224. doi: 10.1105/tpc.106.049585

Shi, M., Luo, X., Ju, G., Yu, X., Hao, X., Huang, Q., et al. (2014). Increased accumulation of the cardio-cerebrovascular disease treatment drug tanshinone in Salvia miltiorrhiza hairy roots by the enzymes 3-hydroxy-3-methylglutaryl CoA reductase and 1-deoxy-D-xylulose 5-phosphate reductoisomerase. *Funct. Integr. Genomics* 14, 603–615. doi: 10.1007/s10142-014-0385-0

Shiu, S. H., and Bleecker, A. B. (2003). Expansion of the receptor-like kinase/Pelle gene family and receptor-like proteins in Arabidopsis. *Plant Physiol.* 132, 530–543. doi: 10.1104/pp.103.021964

Smékalová, V. S., Luptovčiak, I., Komis, G., Šamajová, O., Ovečka, M., Doskočilová, A., et al. (2014). Involvement of YODA and mitogen activated protein kinase 6 in Arabidopsis post-embryogenic root development through auxin up-regulation and cell division plane orientation. *New Phytol.* 203, 1175–1193. doi: 10.1111/nph.12880

Van Gerrewey, T., and Chung, H. S. (2024). MAPK cascades in plant microbiota structure and functioning. *J. Microbiol.* 62, 231–248. doi: 10.1007/s12275-024-00114-3

Wang, C., Wang, G., Zhang, C., Zhu, P., Dai, H., Yu, N., et al. (2017). OsCERK1-mediated chitin perception and immune signaling requires receptor-like cytoplasmic kinase 185 to activate an MAPK cascade in rice. *Mol. Plant* 10 (4), 619–633. doi: 10.1016/j.molp.2017.01.006

Wang, J., Chitsaz, F., Derbyshire, M. K., Gonzales, N. R., Gwadz, M., Lu, S., et al. (2023). The conserved domain database in 2023. *Nucleic Acids Res.* 51, D384–D388. doi: 10.1093/nar/gkac1096

Wang, H., Ngwenyama, N., Liu, Y., Walker, J. C., and Zhang, S. (2007). Stomatal development and patterning are regulated by environmentally responsive mitogen-activated protein kinases in *Arabidopsis*. *Plant Cell* 19, 63–73. doi: 10.1105/tpc.106.048298

Wang, L., Zhao, R., Li, R., Yu, W., Yang, M., Sheng, M., et al. (2018). Enhanced drought tolerance in tomato plants by overexpression of SlMAPK1. *Plant Cell Tissue Organ Culture* 133, 27–38. doi: 10.1007/s11240-017-1358-5

Wang, W., Zheng, Y., Qiu, L., Yang, D., Zhao, Z., Gao, Y., et al. (2024). Genome-wide identification of the SAUR gene family and screening for SmSAURs involved in root development in Salvia miltiorrhiza. *Plant Cell Rep.* 43, 165. doi: 10.1007/s00299-024-03260-5

Wen-Zhi, L., Long-Jiang, Y. U., Wei, L. I., and Hong-Xi, C. (2002). Selecting preparation methods and isolating components of fungal elicitor to enhance Taxol biosynthesis. *Plant Sci. J.* 20, 66–70.

Widmann, C., Gibson, S., Jarpe, M. B., and Johnson, G. L. (1999). Mitogen-activated protein kinase: conservation of a three-kinase module from yeast to human. *Physiol. Rev.* 79 (1), 143–180. doi: 10.1152/physrev.1999.79.1.143

Wu, S. J., Xie, X. G., Feng, K. M., Zhai, X., Ming, Q. L., Qin, L. P., et al. (2022). Transcriptome sequencing and signal transduction for the enhanced tanshinone production in Salvia miltiorrhiza hairy roots induced by Trichoderma atroviride D16 polysaccharide fraction. *Bioscience Biotechnology Biochem.* 86, 1049–1059. doi: 10.1093/bbb/zbac088

Xia, Y., Zhang, L., Hong, X., Huang, Y., Lou, G., Hou, Z., et al. (2023). Metabolomic and antioxidant analyses of Salvia miltiorrhiza Bunge and Salvia prattii Hemsl. seeds. *Natural Product Res.* 16, 1–8. doi: 10.1080/14786419.2023.2269459

Xing, Y., Jia, W., and Zhang, J. (2008). AtMKK1 mediates ABA-induced CAT1 expression and H_2O_2 production via AtMPK6-coupled signaling in *Arabidopsis*. *Plant J.* 54, 440–451. doi: 10.1111/j.1365-313X.2008.03430.x

Xu, Y. H., Wang, J. W., Wang, S., Wang, J. Y., and Chen, X. Y. (2004). Characterization of GaWRKY1, a cotton transcription factor that regulates the sesquiterpene synthase gene (+)-delta-cadinene synthase-A. *Plant Physiol.* 135, 507–515. doi: 10.1104/pp.104.038612

Yamamura, C., Mizutani, E., Okada, K., Nakagawa, H., Fukushima, S., Tanaka, A., et al. (2015). Diterpenoid phytoalexin factor, a bHLH transcription factor, plays a central role in the biosynthesis of diterpenoid phytoalexins in rice. *Plant J.* 84, 1100–1113. doi: 10.1111/tpj.2015.84.issue-6

Zhai, X., Jia, M., Chen, L., Zheng, C. J., Rahman, K., Han, T., et al. (2017). The regulatory mechanism of fungal elicitor-induced secondary metabolite biosynthesis in medical plants. *Crit. Rev. Microbiol.* 43, 238–261. doi: 10.1080/1040841X.2016.1201041

Zhang, Y., Gao, M., Singer, S. D., Fei, Z., Wang, H., and Wang, X. (2012). Genome-wide identification and analysis of the TIFY gene family in grape. *PloS One* 7, e44465. doi: 10.1371/journal.pone.0044465

Zhang, W., and Liu, H. T. (2002). MAPK signal pathways in the regulation of cell proliferation in mammalian cells. *Cell Res.* 12, 9–18. doi: 10.1038/sj.cr.7290105

Zhang, Z., Liu, Y., Huang, H., Gao, M., Wu, D., Kong, Q., et al. (2017). The NLR protein SUMM2 senses the disruption of an immune signaling MAP kinase cascade *via* CRCK3. *EMBO Rep.* 18 (2), 292–302. doi: 10.15252/embr.201642704

Zhao, J., Lou, J., Mou, Y., Li, P., Wu, J., and Zhou, L. (2011). Diterpenoid tanshinones and phenolic acids from cultured hairy roots of Salvia miltiorrhiza Bunge and their antimicrobial activities. *Molecules* 16, 2259–2267. doi: 10.3390/molecules16032259

Zhou, W., Huang, Q., Wu, X., et al. (2017). Comprehensive transcriptome profiling of salvia miltiorrhiza for discovery of genes associated with the biosynthesis of tanshinones and phenolic acids. *Sci. Rep.* 7, 10554. doi: 10.1038/s41598-017-10215-2

Zhou, Y., Singh, S. K., Patra, B., Liu, Y., Pattanaik, S., and Yuan, L. (2025). Mitogen-activated protein kinase-mediated regulation of plant specialized metabolism. *J. Exp. Bot.* 76, 262–276. doi: 10.1093/jxb/erae400

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 20 May 2025
DOI 10.3389/fpls.2025.1581542

OPEN ACCESS

EDITED BY
Simardeep Kaur,
The ICAR Research Complex for North Eastern Hill Region (ICAR RC NEH), India

REVIEWED BY
Subodh Kumar Sinha,
Indian Council of Agricultural Research, India
Sundeep Kumar,
Indian Council of Agricultural Research (ICAR), India

*CORRESPONDENCE
Pradeep Sharma
Pradeep.Sharma@icar.org.in

RECEIVED 22 February 2025
ACCEPTED 21 April 2025
PUBLISHED 20 May 2025

CITATION
Sharma P, Mishra S, Kaur A, Ahlawat OP and Tiwari R (2025) Novel and conserved drought-responsive microRNAs expression analysis in root tissues of wheat (*Triticum asetivum* L.) at reproductive stage. *Front. Plant Sci.* 16:1581542. doi: 10.3389/fpls.2025.1581542

Novel and conserved drought-responsive microRNAs expression analysis in root tissues of wheat (*Triticum asetivum* L.) at reproductive stage

Pradeep Sharma*, Shefali Mishra, Amandeep Kaur, O. P. Ahlawat and Ratan Tiwari

Crop Improvement Division, ICAR-Indian Institute of Wheat and Barley Research, Karnal, India

Introduction: MicroRNAs (miRNAs) are a class of 20- to 24-nucleotide endogenous small RNAs that regulate gene expression post-transcriptionally, playing vital roles in plant development and stress responses. Among abiotic stresses, drought stress (DS) is one of the most critical factors affecting wheat yield worldwide. Understanding miRNA-mediated regulatory mechanisms under drought stress conditions is crucial for improving drought tolerance in wheat.

Methods: To identify drought-responsive miRNAs in wheat, small RNA libraries were constructed from drought-tolerant (NI5439) and drought-susceptible (WL711) genotypes subjected to both control and drought-stress conditions. High-throughput sequencing was used to identify known and novel miRNAs. The family distribution of miRNAs, target prediction, pathway analysis, and differential expression analysis were conducted. A heat map was generated for the top 50 up- and downregulated miRNAs, and novel miRNAs were validated through qRT-PCR.

Results and discussion: A total of 306 known and 58 novel miRNAs were identified across the two wheat genotypes. The identified miRNAs belonged to over 18 families, with miR9662a-3p being the most abundant. Most identified miRNAs were 21 nucleotides in length. A total of 2,300 target genes were predicted for the known miRNAs. Pathway analysis revealed that target genes were involved in key biological processes including signal transduction, transport, organelle localization, DNA methylation, histone and chromatin modification, and plant development. Ten novel miRNAs were validated using qRT-PCR, confirming their differential expression under drought stress. The findings significantly expand the repertoire of drought-responsive and novel miRNAs in wheat. These miRNAs and their target genes provide valuable insights into the molecular mechanisms underlying drought tolerance. The validated novel miRNAs represent potential targets for genetic manipulation to enhance drought resilience in wheat cultivars.

Conclusion: This study provides a comprehensive miRNA expression profile in wheat under drought conditions and highlights several novel miRNAs that are differentially expressed between tolerant and susceptible genotypes. The integration of sequencing, computational analysis, and qRT-PCR validation strengthens the utility of these findings for future functional genomics studies and breeding programs aimed at developing drought-tolerant wheat varieties.

KEYWORDS

wheat, miRNA, transcripts, abiotic stress, drought, DEG, reproductive stage

1 Introduction

Climate change, primarily driven by global warming, has emerged as a significant threat to ecosystems and food security worldwide. Rising temperatures, erratic precipitation patterns, and increased frequency of extreme weather events have intensified challenges in global agriculture, necessitating urgent strategies for climate-resilient food production. Wheat (*Triticum aestivum L.*), one of the most widely cultivated cereal crops, serves as a fundamental source of calories and nutrition for billions of people. However, despite advancements in wheat production over the past decade, global consumption has outpaced supply, exacerbating the demand-supply gap (Parmar et al., 2020; Kaur et al., 2023a; Zhao et al., 2024).

Drought stress at the reproductive stage poses a significant threat to wheat productivity by disrupting key physiological and molecular processes that are essential for grain development. It induces oxidative stress due to excessive accumulation of reactive oxygen species (ROS), leading to cellular damage, impaired photosynthetic efficiency, and premature leaf senescence during the grain-filling stage, ultimately reducing biomass accumulation and grain yield (Nelson et al., 2014; Kim et al., 2020; Berahim et al., 2021; Ahmad et al., 2023). To mitigate these detrimental effects, breeding programs have focused on developing drought-tolerant wheat varieties with enhanced physiological adaptability and stress-responsive molecular mechanisms (Wang et al., 2023). However, the intricate nature of drought responses necessitates a deeper understanding of the regulatory pathways involved in stress adaptation.Phytohormones play a crucial role in modulating wheat's response to drought stress at the reproductive stage. Abscisic acid (ABA) and jasmonic acid (JA) regulate stomatal closure to minimize water loss while activating stress-responsive gene networks through the mitogen-activated protein kinase (MAPK) signaling pathway (Sharma et al., 2023). Additionally, JA influences ABA biosynthesis and degradation, creating a dynamic hormonal interplay that is critical for drought adaptation (Wang et al., 2022). However, excessive accumulation of ABA and JA can accelerate leaf senescence, impairing photosynthetic activity and reducing overall grain yield (Liu et al., 2023). In contrast, brassinosteroids (BRs) contribute to osmotic regulation, while cytokinins (CTKs) delay senescence and counteract the negative effects of ABA and JA, thereby enhancing stress resilience in wheat (Chen et al., 2023; Sun et al., 2023). Understanding these complex hormonal interactions is essential for developing strategies to improve wheat's reproductive-stage drought tolerance and ensure sustainable yield production under water-limited conditions.

Recent advancements in molecular biology have highlighted the role of microRNAs (miRNAs) as key post-transcriptional regulators in plant stress responses. miRNAs are small, non-coding RNA molecules that regulate gene expression by targeting specific mRNAs for degradation or translational repression. Emerging evidence suggests that miRNAs play a crucial role in modulating drought tolerance by influencing stress-responsive pathways such as transcriptional regulation, hormone signaling, and antioxidant defense mechanisms (Sunkar et al., 2012a; Liu et al., 2023). High-throughput sequencing technologies have facilitated the identification of numerous stress-inducible miRNAs in wheat, including miR156, miR166, miR169, miR172, and miR399, which target key genes involved in stress adaptation (Ramachandran et al., 2020; Saroha et al., 2024). Despite growing knowledge of miRNA-mediated stress regulation, the molecular mechanisms underlying miRNA-mRNA interactions in wheat drought responses remain largely unexplored. Functional enrichment analyses and genome-wide expression studies suggest that miRNAs play a significant role in coordinating complex regulatory networks to enhance drought resilience (Mishra et al., 2023a). However, a comprehensive understanding of these regulatory pathways is still lacking, particularly during the critical grain-filling stage when drought stress has the most profound impact on yield formation.

This study aims to elucidate the role of drought-responsive miRNAs in wheat by identifying key miRNA-mRNA regulatory networks involved in drought adaptation. By integrating high-throughput sequencing, transcriptome analysis, and functional validation approaches, we seek to uncover novel miRNA-mediated mechanisms that contribute to drought tolerance in wheat. The findings of this research will provide valuable insights

for breeding climate-resilient wheat varieties with enhanced drought tolerance, ensuring sustainable wheat production in the face of global climate change.

2 Materials and methods

2.1 Stress treatment, tissue collection, and root phenotyping

The study utilized two contrasting wheat genotypes for drought stress, NI5439 as tolerant (T) and WL711 as susceptible (S), to evaluate their performance under control (C) and drought stress (D) conditions (Kaur et al., 2017). Both genotypes were grown in cylindrical mud pots measuring 1.05 m in length and 0.18 m in diameter. The columns were filled with a homogenized mixture of soil, sand, and vermicompost in a 3:1:1 ratio, respectively. Plants in the well-watered treatment were maintained under normal environmental conditions, whereas drought-treated plants were placed in an area covered with a transparent sheet to simulate drought stress. Initially, three germinated seeds were planted per pot, with only one healthy seedling retained after 15 days. Before the onset of drought stress, columns were irrigated twice daily to maintain optimal soil moisture. Drought stress was initiated at the Z24 stage of Zadok's scale (main shoot with four tillers), and root tissues were collected at the Z37 stage (flag leaf just visible). These root samples were immediately flash-frozen in liquid nitrogen and stored at −80°C for subsequent analyses (Zadoks and Board, 1999). The root systems were carefully extracted by breaking the pots and sectioned into four depths: 0–30 cm, 30–60 cm, 60–90 cm, and 90–120 cm, following the approach of Narayanan et al. (2014). Roots were gently washed using a low-pressure water fountain over a 1.5 m sieve to minimize damage. The cleaned root samples were then preserved in 70% ethanol and scanned using a document scanner. Root volume and other traits like root length and diameter were quantified using WinRHIZO® software, which provides an accurate digital analysis of root morphological parameters (Singh et al., 2011).

2.2 RNA extraction, construction of small RNA libraries and deep sequencing

Total RNA was extracted using the TRIzol method, and its quality was assessed via NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, USA).Small RNAs were isolated in triplicate from frozen root tissues using the mirVana™ miRNA isolation kit (Ambion), following the manufacturer's protocol. The small RNA fractions from the three replicates were pooled together for library construction. The preparation of small RNA libraries was performed according to the protocol by (Lu et al., 2007), with minor modifications. Small RNAs were sequentially ligated with 3' and 5' adapters (Supplementary Table 1). Following this, reverse transcription was performed using an RT primer, and PCR amplified the resulting cDNA. The integrity and quantity of the constructed libraries were assessed using the RNA Integrity Number (RIN) on an Agilent 2100 Bioanalyzer (Agilent Technologies, USA). The final libraries were submitted for high-throughput sequencing at SciGenome Labs (India) using the Illumina MiSeq platform (Illumina, USA). The libraries were labelled as Tolerant Control (TC), Tolerant Drought (TD), Susceptible Control (SC) and Susceptible Drought (SD). The raw data have been submitted to NCBI and the accession number is PRJNA1012115.

2.3 Computational analysis of small RNA sequencing data

Sequencing reads with a Phred quality score > 30 were retained and processed using Cutadapt v1.3 to remove adapter sequences. Non-coding small RNAs, including siRNAs, snRNAs, snoRNAs, piRNAs, tRNAs, and rRNAs, were filtered out by mapping them to respective databases (Supplementary Table 2) using Bowtie2 v2.1.0 (Langmead and Salzberg, 2012). The remaining non-redundant reads (17–35 nucleotides) were used for identifying both conserved and novel miRNAs in wheat. For conserved miRNAs, reads were aligned to the Triticum aestivum reference genome using Bowtie v1.2.3 (Langmead et al., 2009) and compared with known miRNAs in miRBase v22 (Kozomara et al., 2019). Initially, reads were mapped to mature miRNAs, followed by precursor sequences to ensure accuracy. Differential expression analysis was performed using DESeq2 in R v4.0.0 (Love et al., 2014) to identify significantly differentially expressed miRNAs.

For novel miRNA identification, miRDeep2 v2.0.0.7 (Friedländer et al., 2012) was used with the *Triticum aestivum* genome as a reference (Ensembl Plants Release 60: https://ftp.ebi.ac.uk/ensemblgenomes/pub/release-60/plants/fasta/triticum_aestivum/ncrna/). High-confidence novel miRNAs were identified based on (Meyers et al., 2008) criteria, including a 3' two-nucleotide overhang, no more than four mismatches with the complementary precursor arm, and minimal asymmetric bulges (one or two bases) in the miRNA/miRNA* duplex. Secondary structure prediction was performed using MFOLD (Zuker, 2003) to confirm the characteristic stem-loop hairpin structure of miRNA precursors. Potential target genes for novel miRNAs were predicted using miRanda, which identifies mRNA targets based on sequence complementarity. Thisworkflow, integrating updated genome references and bioinformatics tools, ensures high-precision miRNA discovery and expression analysis, providing insights into their regulatory roles in gene silencing and stress response pathways in wheat.

2.4 Gene ontology analysis

To understand the functional roles of differentially expressed genes (DEGs) in wheat under drought stress, Gene Ontology (GO)

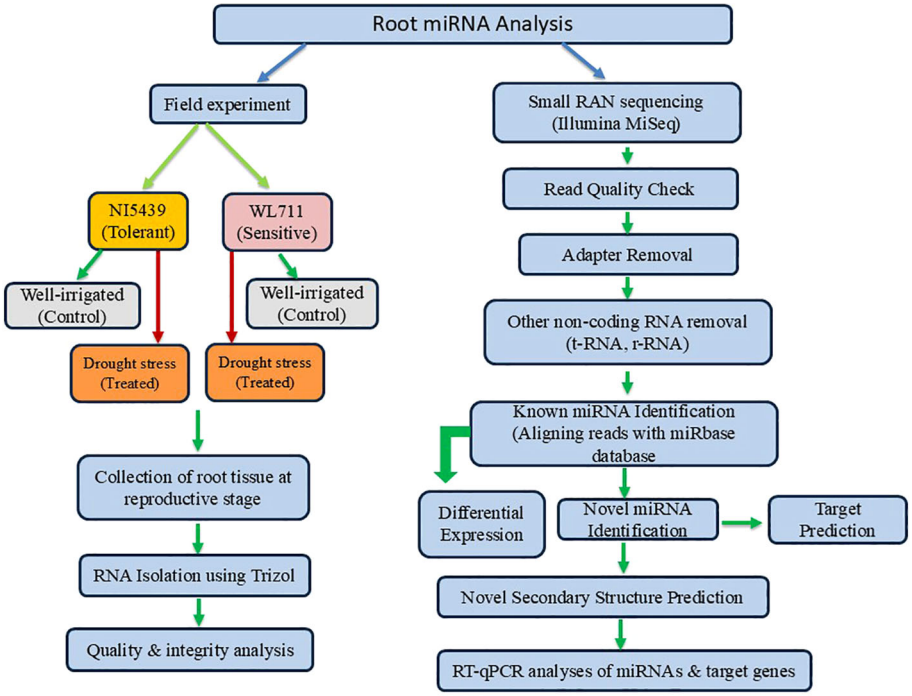


FIGURE 1
Summary of the approach employed for the identification, characterization, and functional validation of known and novel drought-responsive miRNAs in wheat root tissues at the reproductive stage.

enrichment analysis was performed. The DEGs were annotated using the BLAST2GO tool (Conesa et al., 2005) by mapping sequences against the non-redundant (NR) database of the National Center for Biotechnology Information (NCBI). The identified genes were classified into three major GO categories: biological process (BP), molecular function (MF), and cellular component (CC). The functional enrichment analysis was conducted using AgriGO v2.0, a specialized tool for plant GO analysis (Tian et al., 2017), to determine significantly overrepresented GO terms. The statistical significance of enrichment was assessed using Fisher's exact test with a false discovery rate (FDR) < 0.05 for multiple testing corrections. Figure 1 presents a summary of the miRNA analysis conducted under stress conditions, highlighting the key findings and insights from the study.

2.5 Gene regulatory network analysis

Cytoscape (version 3.2.1)tool was used for analysis of gene network analysis of differential expressed genes. For network analysis, top 100 upregulated and downregulated genes each were considered. ARACNE (Algorithm for the Reconstruction of Accurate Cellular Networks) and Network Analyzer plug-in were used for analyzing the network of all the four sets of DEGs. On the basis of high degree and betweenness, hub genes were selected.

2.6 Validation and expression profiling of miRNAs and their target genes by quantitative PCR

Small RNA cDNA (srcDNA) library was constructed according to the protocol of (Bailey et al., 2013). Briefly, small RNAs were polyadenylated at 37°C for 45 min in 50 µl reaction volume containing 0.3 µg ofmall RNA, 0.1 U *E. coli* poly(A) polymerase, 1X *E. coli* poly(A) polymerase reaction buffer [50 mM Tris-HCl, 250 mM NaCl, 10 mM $MgCl_2$, pH 7.9 at 25°C] and 1 mM ATP. Then, the poly (A)-tailed small RNA samples were purified to remove unincorporated ATP by using a purification cartridge provided in mirVana probe and marker kit as per manufacturer's protocols. The purified poly(A)-tailed small RNA samples were stored at –70°C. The srcDNA libraries were generated by mixing 500 ng of poly(A)-tailed small RNA and 1 µg of RTQ primer in a 26 µl reaction volume. The reaction mixture was incubated at 65°C for 10 min followed by addition of 0.2 U M-MuLV reverse transcriptase, 1XM-MuLV reverse transcriptase reaction buffer [50 mM Tris-HCl, 75 mM KCl, 3 mM $MgCl_2$, 10 mM DTT, pH 8.3 at 25°C] and 1 mM dNTP mix in a final reaction volume of 40 µl. The reverse transcription was carried out at 37°C for 60 min followed by inactivation of the enzyme at 70°C for 15 min. About 1 µl of 5 U RNaseH was added to remove poly(A)-tailed small RNAs. The samples were purified by using the QIAquick PCR purification kit in 50 µl of final volume.

The qRT-PCR was performed using 0.3 µg of srcDNA, 1X SYBR green/fluorescein qPCR master mix, 1 µM RTQ-UNIr primer, and 1 µM miRNA-specific forward primer (Supplementary Table 3). The PCR reactions were performed in triplicate for each gene. The thermal cycling PCR reactions were performed with the following profile: 95°C for 5 min, 40 cycles of 15 sec denaturing at 94°C, 30-sec annealing at 55°C and 30-sec extension at 72°C, and finally a melt curve step from 65°C to 95°C with a rise of 0.5°C for 5 sec. The U6 snRNA was used as a reference gene for all the samples amplified. Relative quantification of expression for each miRNA was analyzed using the comparative CT method as described by (Livak and Schmittgen, 2001). Ten novel miRNAs viz. #ps_55, #ps _199, #ps _45 and #ps_160, #ps_19, #ps_91, #ps_187, #ps_103, #ps_74, #ps_47, #ps_89, #ps_157, #ps_55, #ps_121 and its targets TaDRA1, TaDRA2, TaDRA3, TaDRA4, TaDRA5, TaDRA6, TaDRA7, TaDRA8, TaDRA9, TaDRA10, TaDRA11, TaDRA12, TaDRA13 and TaDRA14 were chosen at random to validate them under drought stress in wheat respectively. The real ids of miRNAs and its targets were mentioned in (Supplementary Table 4).

3 Results

3.1 Phenotyping under drought stress

Two wheat genotypes that were already known for their distinct behavior in drought stress were grown in fields under irrigated and drought conditions. Root samples were collected for small RNA library preparation at the booting stage of growth and development. The pooled analysis of variance revealed that both genotype and stress conditions differed significantly for all the traits studied. Based on LSD, the two genotypes differed in root length, surface area, and Length perVolume whereas stress conditions were significantly different for root length, surface area, and root volume.The two genotypes under normal conditions differed significantly for root length, surface area, and Length perVolume (Figure 2). However, under drought stress conditions, the two genotypes differed in total length and Length perVolume.Genotype 2 registered more reduction than genotype 1 in all the traits except diameter (Figure 2). Under normal conditions, the root length had a significant positive correlation with surface area and Length perVolume and the root volume had a significant positive correlation with surface area and average diameter. However, under drought-stressconditions, all the traits were interlinked positively.

3.2 Analysis of the small RNA libraries

Small RNA libraries were generated from the root tissues of two wheat genotypes, one exhibiting drought tolerance and the other drought sensitivity, cultivated under both normal and drought-stressed conditions. Sequencing produced approximately 200 million raw reads (200,780,257). Specifically, 62,320,791 reads were obtained from the drought-tolerant genotype under normal conditions, 73,006,712 from the same genotype under drought stress, 24,620,142 from the drought-sensitive genotype under normal conditions, and 40,832,612 from the drought-sensitive genotype under drought stress (Table 1). Raw sequencing reads underwent preprocessing to remove adapter sequences, yielding clean, non-redundant reads of 10,825,961, 6,203,540, 4,552,914, and 8,142,717 for the respective libraries. Non-coding RNAs, including siRNAs, snRNAs, snoRNAs, piRNAs, tRNAs, and rRNAs, were filtered out, leaving unannotated reads ranging from 17 to 35 nucleotides in length. These unannotated sequences were then analyzed to identify both conserved and novel miRNAs in wheat, offering valuable insights into miRNA-based regulatory mechanisms in response to drought stress.

3.3 Identification of conserved miRNAs

The identification of conserved miRNAs was conducted by aligning 17–35 bp unique reads to a wheat reference genome. Initially, these reads were mapped to mature miRNAs, and any unmapped sequences were subsequently aligned with precursor sequences. This approach led to the identification of 150, 132, 146, and 167 previously characterized miRNAs, distributed across 86 families, in the tolerant-control (TC), tolerant-stressed (TD), sensitive-control (SC), and sensitive-stressed (SD) libraries, respectively (Table 2). Among these, 119 miRNAs were found to be shared across all libraries analyzed in this study (Figure 3). Additionally, several miRNAs were uniquely expressed under individual conditions, such as 8 in TC, 3 in SC, 15 in SD, and 1 in TD—suggesting condition-specific regulatory roles. Shared miRNAs between pairs and triplets of conditions also varied, with notable overlap between SC and SD (10 miRNAs) and between SC and TC (9 miRNAs). These results highlight both common and condition-specific miRNA-mediated regulatory responses under the tested stress conditions.

Notably, two miRNAs (tae-miR1119 and tae-miR9773) were exclusively detected under control conditions in both genotypes, whereas three miRNAs (tae-miR1129, tae-miR9660, and tae-miR9661) were specifically recovered under drought stress. These five miRNAs are likely to play a regulatory role in drought stress response. Additionally, 12 genotype-specific miRNAs were identified, with two (tae-miR1125 and tae-miR5049) associated with the drought-tolerant genotype, while the remaining ten (tae-miR1136, tae-miR9652, tae-miR9657b, tae-miR9657c, tae-miR9663, tae-miR9666a, and tae-miR9670) were linked to the drought-sensitive genotype.

The number of miRNA members varied across different families. The tae-miR159 and tae-miR9662 families exhibited the highest diversity, with an average of 20 members each. These were followed by tae-miR9653, tae-miR167, tae-miR1130, tae-miR9672, and tae-miR9657, each containing more than 10 members. The most highly conserved miRNA identified across all wheat libraries was tae-miR9662a-3p, along with its precursor tae-MIR9662a (Supplementary Table 5). The second and third most abundant miRNAs belonged to the conserved MIR159 family, specifically tae-miR159a and tae-miR159b. In terms of length distribution, the most prevalent class of conserved miRNAs was 21 nucleotides (28.57% on average), followed by the 19-nucleotide class (15.63% on average) (Figure 4).

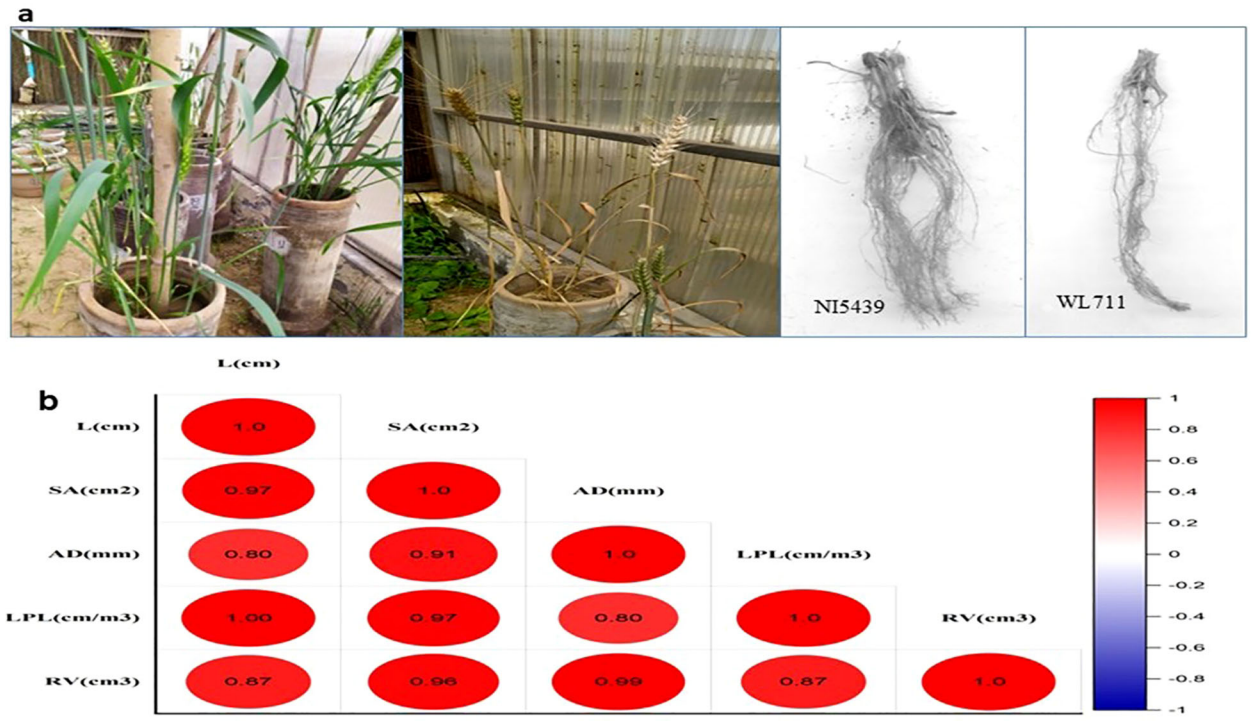


FIGURE 2
(A) Phenotyping of contrasting wheat genotypes under drought stress conditions to evaluate variations in root traits. **(B)** Correlation Matrix of Root Morphological Traits in Wheat under Drought Stress. L (cm) – Root Length, SA (cm^2) – Root Surface Area, AD (mm) – Root Average Diameter, LPL (cm/m^3) – Root Length Per Unit Volume, RV (cm^3) – Root Volume. The figure represents a Pearson correlation matrix for root traits in wheat, with color intensity indicating correlation strength (red = positive, blue = negative).

3.4 Identification of Novel miRNAs

The discovery of novel miRNAs was achieved through secondary structure prediction, leading to the identification of 58 novel miRNAs in *Triticum aestivum*. These miRNAs were predicted from aligned sequencing data based on structural and genomic features. Among them, 9 novel miRNAs were detected in the control condition of the NI5439 genotype, while 10 were identified under drought stress. Similarly, in the WL711 genotype, 15 and 14 novel miRNAs were detected in control and drought-stressed samples, respectively (Table 3). Comprehensive characterization was conducted for each novel miRNA, including chromosomal localization, precursor sequence identification, secondary structure prediction, and mature miRNA sequence analysis. The secondary stem-loop structures were computationally predicted using an energy minimization approach. The minimal folding free energy (MFE) values of precursor miRNAs ranged from –46.71 to –13.47 kcal/mol, with an average of –33.73 ± 10.57 kcal/mol (Table 3), indicating their structural stability. The graphical representation of these structures is provided in Supplementary Figure 1a-d.

Interestingly, two novel miRNAs, #ps19 and #ps91, were consistently detected in both libraries of the drought-sensitive genotype, suggesting their potential involvement in drought stress regulation. The most prevalent length among the novel miRNAs was 21 nucleotides, with a range spanning from 18 to 23 nucleotides. The GC content of these novel miRNAs averaged 55.67 ± 13.16%, indicative of their potential stability and functional relevance. Regarding chromosomal distribution, 13 loci were mapped to genome A, whereas genome B and genome D harbored 26 and 20 loci, respectively. Notably, the highest number of miRNA loci detected on a single chromosome was 8, found on chromosomes 5B and 5D, highlighting possible regions of miRNA enrichment in the wheat genome. These findings provide new insights into the regulatory landscape of wheat miRNAs, particularly under drought stress conditions.

3.5 Gene ontology

To annotate and analyze the functional roles of predicted miRNA target genes in wheat, a total of 4,551 target transcripts (740 in TC, 894 in TD, 1,340 in SC, and 1,577 in SD) were subjected to Gene Ontology (GO) analysis. These transcripts, associated with genes of known functions, were classified into biological processes, cellular components, and molecular functions based on their GO annotations (Figure 5).

In the molecular function category, the predicted miRNA targets were primarily linked to DNA, nucleic acid, and ion binding (183 terms in TC, 167 in TD, 302 in SC, and 322 in SD), catalytic activity (20 in TC, 13 in TD, 47 in SC, and 40 in SD), transferase activity (24 in TC, 21 in TD, 64 in SC, and 36 in

10.3389/fpls.2025.1581542

TABLE 1 Summary of small RNA sequencing data in the four libraries ofwheat.

Category	TC (Tolerant-Control)			TD(Tolerant-Stressed)			SC (Sensitive-Control)			SD (Sensitive-Stressed)		
	Total Reads	Percent (%) Total Reads	Unique Reads	Total Reads	Percent (%) Total Reads	Unique Reads	Total Reads	Percent (%) Total Reads	Unique Reads	Total Reads	Percent (%) Total Reads	Unique Reads
Total raw reads	6,23,20,791	–	–	7,30,06,712	–	–	2,46,20,142	–	–	4,08,32,612	–	–
Clean reads	1,08,25,961	100	–	62,03,540	100	–	45,52,914	100	–	81,42,717	100	–
siRNA	1,33,821	1.24	627	1,13,169	1.82	627	97,455	2.14	627	1,43,269	1.76	627
piRNA	8,56,115	7.91	1,66,339	6,71,632	10.83	1,55,552	5,90,126	12.96	1,45,855	9,81,213	12.05	1,66,417
snRNA	7,696	0.07	289	6,926	0.11	287	4,687	0.10	286	6,700	0.08	292
snoRNA	15,133	0.14	1,060	13,249	0.21	1,049	9,954	0.22	1,032	15,547	0.19	1,058
tRNA	3,36,752	3.11	51,000	1,95,055	3.14	42,543	1,81,407	3.98	41,918	2,82,927	3.47	50,266
rRNA	40,97,899	37.85	9,31,773	21,57,834	34.78	6,30,847	17,59,917	38.65	5,80,263	26,39,542	32.42	7,92,532
Unannotated reads (4bp – 50bp)	53,78,545	49.68	–	30,45,675	49.10	–	19,09,368	41.94	–	40,73,519	50.03	–

TABLE 2 Summary of conserved and novel miRNA in wheat under drought stress.

Sample Name	TC	TD	SC	SD
No. of mature miRNAs	59	52	60	72
No. of precursormiRNAs	91	80	86	95
No. of novel miRNAs	9	17	18	15

SD), protein kinase activity (26 in TC, 97 in TD, 136 in SC, and 115 in SD), and oxidoreductase activity (20 in TC, 22 in TD, 70 in SC, and 36 in SD) (Figure 5; Supplementary Table 6a-d). In the biological processes category, many target transcripts were associated with stress response and defense mechanisms (17 terms in TC, 27 in TD, 50 in SC, and 35 in SD) and regulatory functions, including metabolic processes, growth and development, signal transduction, transcriptional regulation, and photosynthesis (Figure 5; Supplementary Table 6a-d). Similar patterns have been observed in rice and maize (Sunkar et al., 2012b; Božić et al., 2024). The cellular component category revealed that most target genes were associated with membrane components, while fewer were linked to organelles such as chloroplasts, mitochondria, ribosomes, and spliceosomes. The lower representation of chloroplast-related genes suggests a suppression of photosynthetic activity under stress, supported by the decline in photosynthesis-related GO terms in the biological process category. Interestingly, in resistant plants, plastid-associated terms increased under the cellular component category, although photosynthetic activity remained suppressed.

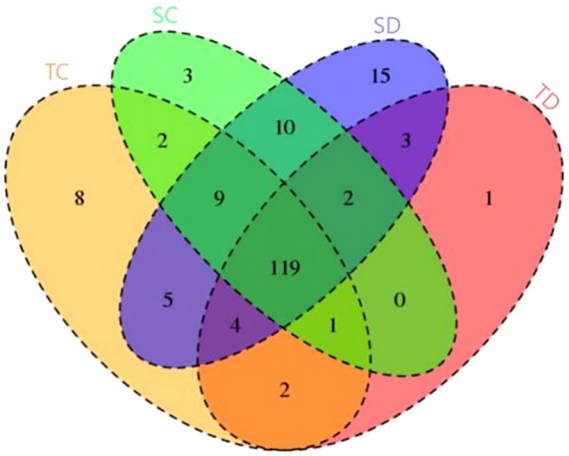


FIGURE 3
Venn diagram represents distribution of conserved miRNA among four libraries. illustrates the distribution of differentially expressed genes (DEGs) across four experimental conditions: TC (yellow), SC (green), SD (blue), and TD (red). Each section represents unique or shared DEGs among the conditions. The numbers indicate the count of DEGs specific to one condition or shared among multiple conditions.

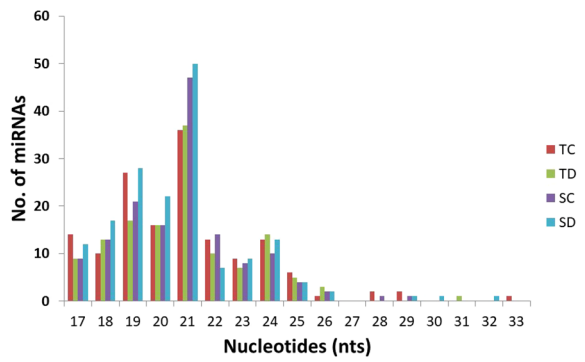


FIGURE 4
Length distribution of conserved miRNAs in the four libraries from wheat. TC: sRNA library from the drought-tolerant genotype NI-5439 without drought stress, TD: from NI-5439 with drought stress, SC: from the drought-sensitive genotype WL-711 without drought stress, SD: from WL-711 with drought stress.

3.6 Gene regulating network of miRNAs

The identification of miRNA-regulated target genes is essential for elucidating the functionalroles of miRNAs in plants, particularly in response to environmental stress. Computational target prediction revealed 314 putative target genes for differentially expressed miRNAs in wheat, spanning various biological and molecular functions. In the tolerant-control (TC) library, five miRNAs were predicted to regulate 41 genes, while 13 novel miRNAs were associated with the regulation of 60 genes in the tolerant-drought (TD) library (Figure 6). Similarly, in the sensitive-control (SC) library, 11 miRNAs were linked to 78 target genes, whereas in the sensitive-drought (SD) library, eight novel miRNAs were predicted to influence the expression of 134 genes under both control and drought stress conditions (Supplementary Figure 2a-c).

The predicted target genes exhibit a broad spectrum of functional categories, underscoring their pivotal roles in plant

TABLE 3 Novel miRNAs identified by reference genome of *Triticum aestivum* in NI5439 and WL711 samples under control and drought stress condition.

miRNA ID	MFE	Chromosome	Sequence	Mature miRNA length
(i) NI5439 control				
#ps111	-38.88	2A	uggacgaggaugugcagcugc	21
#ps141	-38.88	2B	uggacgaggaugugcagcugc	21
#ps25	-43.58	2D	uggacgaggaugugcagcugc	21
#ps45	-18.89	3A	gcuugggcgagaguaguacuagg	23
#ps55	-42.05	5B	ugaagcugccagcaugaucuga	22
#ps59	-46.71	5B	ucggaccaggcuucauucccc	21
#ps87	-41.69	5D	ugaagcugccagcaugaucuga	22
#ps93	-46.58	5D	ucggaccaggcuucauucccc	21
#ps56	-42.72	6D	cccgccuugcaccaagugaa	20
(ii) NI5439 drought stress				
#ps187	-38.88	2A	uggacgaggaugugcagcugc	21
#ps172	-13.47	2A	gaagacugcucugcuuugag	20
#ps265	-38.88	2B	uggacgaggaugugcagcugc	21
#ps47	-43.58	2D	uggacgaggaugugcagcugc	21
#ps103	-46.71	5B	ucggaccaggcuucauucccc	21
#ps157	-46.58	5D	ucggaccaggcuucauucccc	21
#ps22	-22.31	6A	aaagacugcucugcuuugag	20
#ps39	-22.31	6B	aaagacugcucugcuuugag	20
#ps74	-42.72	6D	cccgccuugcaccaagugaa	20
#ps191	-21.9	7B	ugauuguugcuugcguacacu	21

(Continued)

TABLE 3 Continued

miRNA ID	MFE	Chromosome	Sequence	Mature miRNA length
(iii) WL711 control				
#ps137	-38.88	2A	uggacgaggaugugcagcugc	21
#ps145	-30.74	2A	ccucgccggcugcgcguccacc	22
#ps185	-38.88	2B	uggacgaggaugugcagcugc	21
#ps19	-43.58	2D	uggacgaggaugugcagcugc	21
#ps34	-21.3	3A	ugcugcguugacuggcgcuc	20
#ps115	-32.46	3B	ccucgccggcugcgcguccacc	22
#ps91	-46.71	5B	ucggaccaggcuucauuccccc	21
#ps94	-32.93	5B	ccucgccggcugcgcguccacc	22
#ps87	-42.05	5B	ugaagcugccagcaugaucuga	22
#ps112	-32.93	5D	ccucgccggcugcgcguccacc	22
#ps125	-41.69	5D	ugaagcugccagcaugaucuga	22
#ps133	-46.58	5D	ucggaccaggcuucauuccccc	21
#ps43	-15	6B	uauauuaucacucugaggga	20
#ps161	-21.9	7B	ugauuguugcuugcguacacu	21
#ps63	-42.72	6D	ucgcuuggugcagaucgggac	21
(iv) WL711 drought stress				
#ps54	-34.27	1B	ugagaagguagaucauaauagc	22
#ps55	-25.65	1B	uguuaugaucugcuucucauc	20
#ps121	-38.88	2A	uggacgaggaugugcagcugc	21
#ps157	-38.88	2B	uggacgaggaugugcagcugc	21
#ps19	-43.58	2D	uggacgaggaugugcagcugc	21
#ps89	-42.05	5B	ugaagcugccagcaugaucuga	21
#ps91	-46.71	5B	ucggaccaggcuucauuccccc	21
#ps123	-41.69	5D	ugaagcugccagcaugaucuga	22
#ps131	-46.58	5D	ucggaccaggcuucauuccccc	21
#ps20	-22.31	6A	aaagacugcucugcuuugag	
#ps25	-22.31	6B	aaagacugcucugcuuugag	20
#ps55	-42.72	6D	ucgcuuggugcagaucgggac	21
#ps168	-41.48	7D	ugcaucauuuggaacucgccg	20
#ps173	-33.29	7D	uuccaaguugcguaguggaccgg	23

growth, development, and stress adaptation. Notably, several identified miRNAs target key transcription factor (TF) families, including MYB, NAC, WRKY, and bZIP, which serve as master regulators of gene expression networks balancing abiotic stress responses, hormone signaling, and developmental processes. The enrichment of TFs among miRNA targets suggests a hierarchical regulatory mechanism wherein miRNAs modulate multiple downstream genes, thereby influencing extensive biological pathways. Additionally, a subset of target genes encodes proteins associated with phytohormonal signaling, such as auxin response factors (ARFs) and abscisic acid (ABA)-responsive elements, which are integral to stress tolerance, growth regulation, and developmental plasticity under drought stress. The identification of histone-modifying enzymes, including histone deacetylases (HDACs) and methyltransferases, implies a potential role for miRNAs in chromatin remodeling and epigenetic regulation, thereby contributing to transcriptional stability and adaptive gene expression in response to environmental stressors. Furthermore, several miRNAs were predicted to regulate genes involved in ion transport and homeostasis, such as potassium and calcium

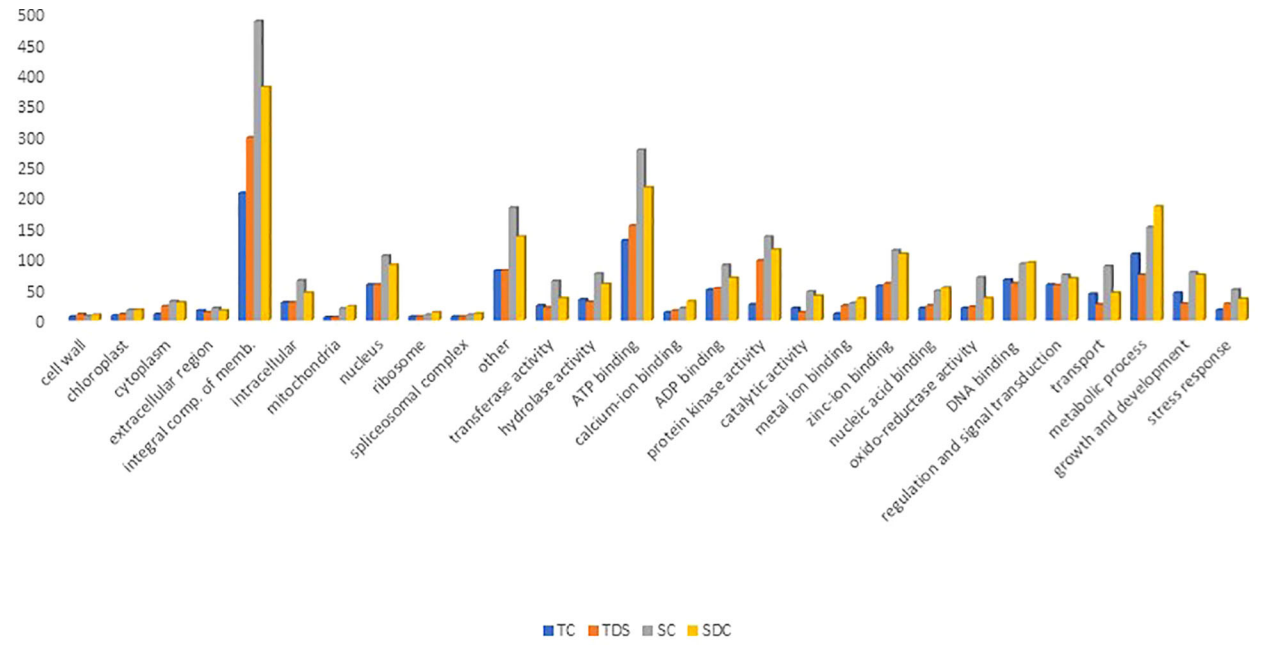


FIGURE 5
Gene ontology analysis of pooled libraries under control and stress conditions.

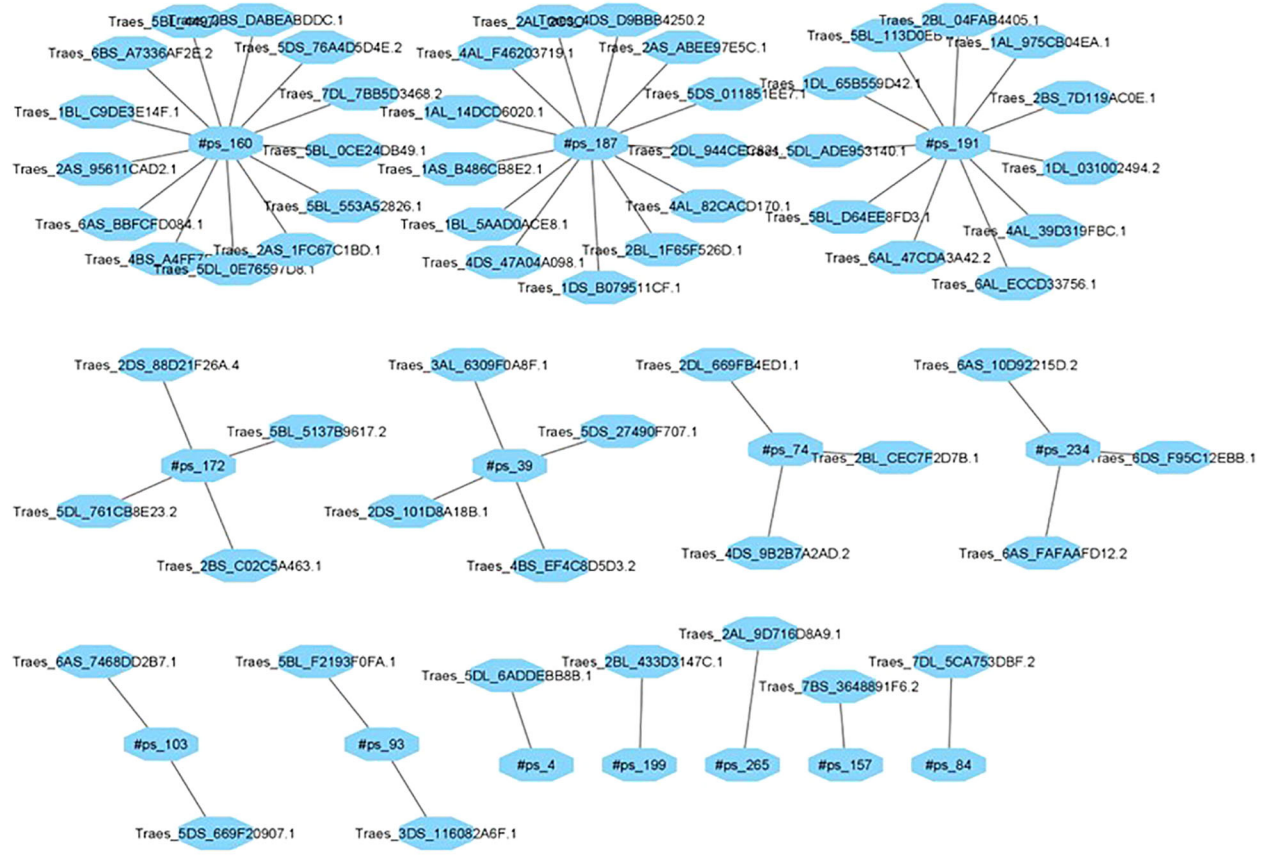


FIGURE 6
Network of miRNA targets in wheat genes under drought stress in the TD library.

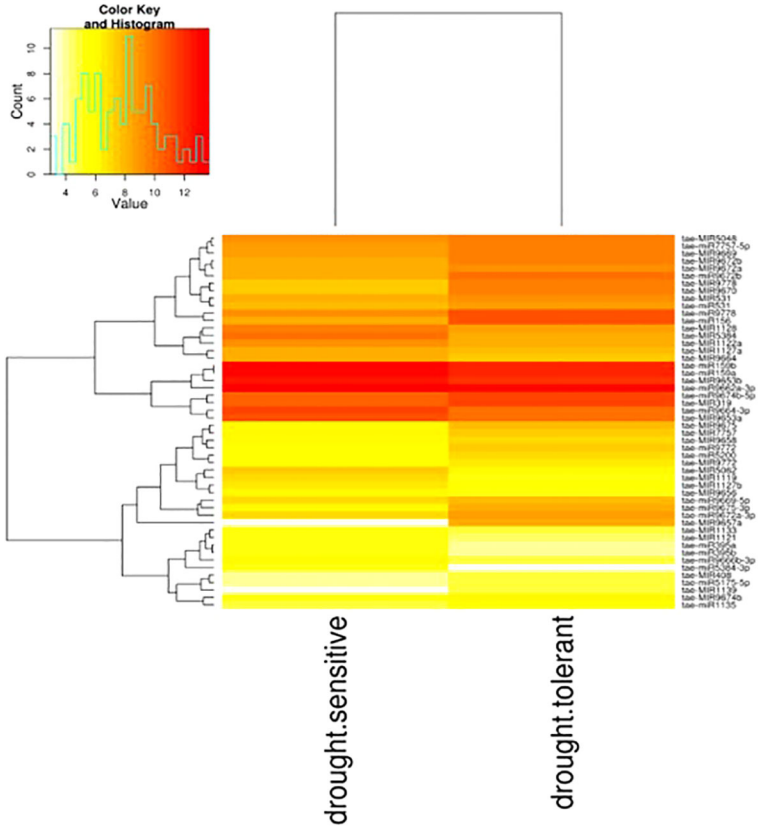


FIGURE 7
Visualizing differential expression between sample drought-tolerant and drought-sensitive using g-plot heatmap from Wheat. The plot shows the hierarchical clustering of sample both samples. (Yellow shows low-level expression; Red shows high-level expression).

transporters, which are essential for maintaining osmotic balance and intracellular signaling under drought conditions. The identification of miRNA-targeted proteases suggests a role in stress-induced proteolysis, which may facilitate cellular adaptation by regulating protein turnover under adverse conditions. Additionally, the modulation of kinase-mediated phosphorylation cascades by miRNAs indicates their involvement in signal transduction pathways, influencing cellular responses to external stimuli. Collectively, these findings highlight the intricate miRNA-mediated regulatory networks that fine-tune gene expression, thereby enhancing wheat's adaptive potential to drought stress. A deeper understanding of these interactions could provide a foundation for developing stress-resilient wheat cultivars through targeted genetic and biotechnological approaches.

3.7 *In silico* expression analysis of miRNAs

The differential expression patterns of conserved miRNAs in both wheat genotypes were analyzed to assess their regulatory roles under drought stress conditions. Hierarchical clustering of all samples was performed to visualize expression trends, with the heatmap representing distinct expression profiles across different experimental conditions (Figure 7).

A total of 18 mature miRNAs exhibited significant differential expression based on fold change criteria (<1 or >1), indicating their potential involvement in drought stress response. Among them, six miRNAs—*tae-miR395a, tae-miR395b, tae-miR5049-3p, tae-miR5384-3p, tae-miR9664-3p*, and *tae-miR9666b-3p*—were significantly upregulated during drought stress, suggesting their potential role in activating stress-responsive pathways (Table 4). These miRNAs may be involved in regulating sulfur metabolism, oxidative stress responses, and transcriptional regulation under water-limiting conditions.

Conversely, two-thirds of the differentially expressed miRNAs were downregulated under drought stress, including *tae-miR156, tae-miR1135, tae-miR531, tae-miR5175-5p, tae-miR5200, tae-miR7757-5p, tae-miR9669-5p, tae-miR9672a-3p, tae-miR9672b, tae-miR9675-3p, tae-miR9772*, and *tae-miR9778*. The repression of these miRNAs may be linked to the modulation of stress-adaptive processes such as leaf morphogenesis, hormone signaling, and secondary metabolite biosynthesis, which are critical for drought tolerance.These results suggest that specific miRNAs play contrasting roles in drought stress adaptation by either enhancing or suppressing gene expression networks associated with plant survival under water-deficient conditions. The observed differential expression patterns provide valuable insights into the complex regulatory mechanisms governing stress

10.3389/fpls.2025.1581542

TABLE 4 Comparative expression Profile of miRNAs in contrasting wheat genotypes.

S.No.	miRNA	log2	pval	padj
1.	tae-miR156	-3.15	0.02	0.63
2.	tae-miR9778	-2.82	0.03	0.64
3.	tae-miR9675-3p	-2.67	0.05	0.77
4.	tae-miR9672a-3p	-2.54	0.05	0.77
5.	tae-miR9672b	-2.29	0.07	0.82
6.	tae-miR9772	-2.07	0.07	0.82
7.	tae-miR531	-1.48	0.24	1.00
8.	tae-miR5200	-1.44	0.15	1.00
9.	tae-miR9669-5p	-1.25	0.34	1.00
10.	tae-miR5175-5p	-1.25	0.21	1.00
11.	tae-miR7757-5p	-1.04	0.40	1.00
12.	tae-miR1135	-1.04	0.28	1.00
13.	tae-miR9666b-3p	1.06	0.28	1.00
14.	tae-miR395b	1.18	0.24	1.00
15.	tae-miR5049-3p	1.18	0.51	1.00
16.	tae-miR395a	1.28	0.20	1.00
17.	tae-miR9664-3p	1.65	0.15	1.00
18.	tae-miR5384-3p	2.51	0.02	0.63

responses in wheat, offering potential targets for improving drought resilience through genetic or biotechnological approaches.

3.8 Validation of novel miRNAs

High-throughput deep sequencing of root tissues from contrasting wheat genotypes revealed several putative novel miRNAs involved in drought stress response. To validate these findings, plants were cultivated under both control and drought conditions, followed by qPCR-based expression analysis in root tissues. A subset of ten novel miRNAs and their corresponding target genes were randomly selected for validation, including *#ps_55, #ps_199, #ps_45, #ps_160, #ps_19, #ps_91, #ps_187, #ps_103, #ps_74, #ps_47, #ps_89, #ps_157, #ps_55*, and *#ps_121*. Their predicted target genes (*TaDRA1–TaDRA14*) encode proteins involved in drought response and adaptation mechanisms, including stress-responsive transcription factors, osmotic regulation proteins, and key enzymes in antioxidant defense pathways.

Expression profiling in root tissues revealed genotype-specific regulatory patterns of the novel miRNAs, with distinct expression dynamics between the drought-tolerant (NI5439) and drought-sensitive (WL711) genotypes (Figures 8A, B). Notably, *#ps_199* exhibited significantly higher expression in the root tissues of the NI5439 genotype compared to other miRNAs, suggesting a potential role in root-specific drought response mechanisms. The NI5439 genotype displayed elevated expression levels of novel miRNAs relative to WL711, indicating a more robust miRNA-mediated regulatory network in drought adaptation.The identified miRNAs and their associated target genes likely participate in key physiological and molecular pathways, including hormone signaling, osmoprotectant biosynthesis, oxidative stress mitigation, and root system architecture modulation.

4 Discussion

Drought stress is one of the most significant abiotic factors limiting wheat productivity (*Triticum aestivum* L.). To cope with water-deficit conditions, plants have evolved complex regulatory mechanisms, including miRNA-mediated gene regulation, which plays a crucial role in modulating stress responses at the post-transcriptional level. This study provides a comprehensive analysis of drought-responsive miRNAs in wheat, identifying both conserved and novel miRNAs, along with their target genes and functional pathways. The findings were validated through expression profiling, functional annotation, and comparative analysis with prior studies, offering insights into miRNA-mediated drought tolerance mechanisms.

The identification of conserved and novel miRNAs is crucial for understanding the regulatory landscape of drought stress responses. In this study, multiple conserved miRNAs, including tae-miR159,

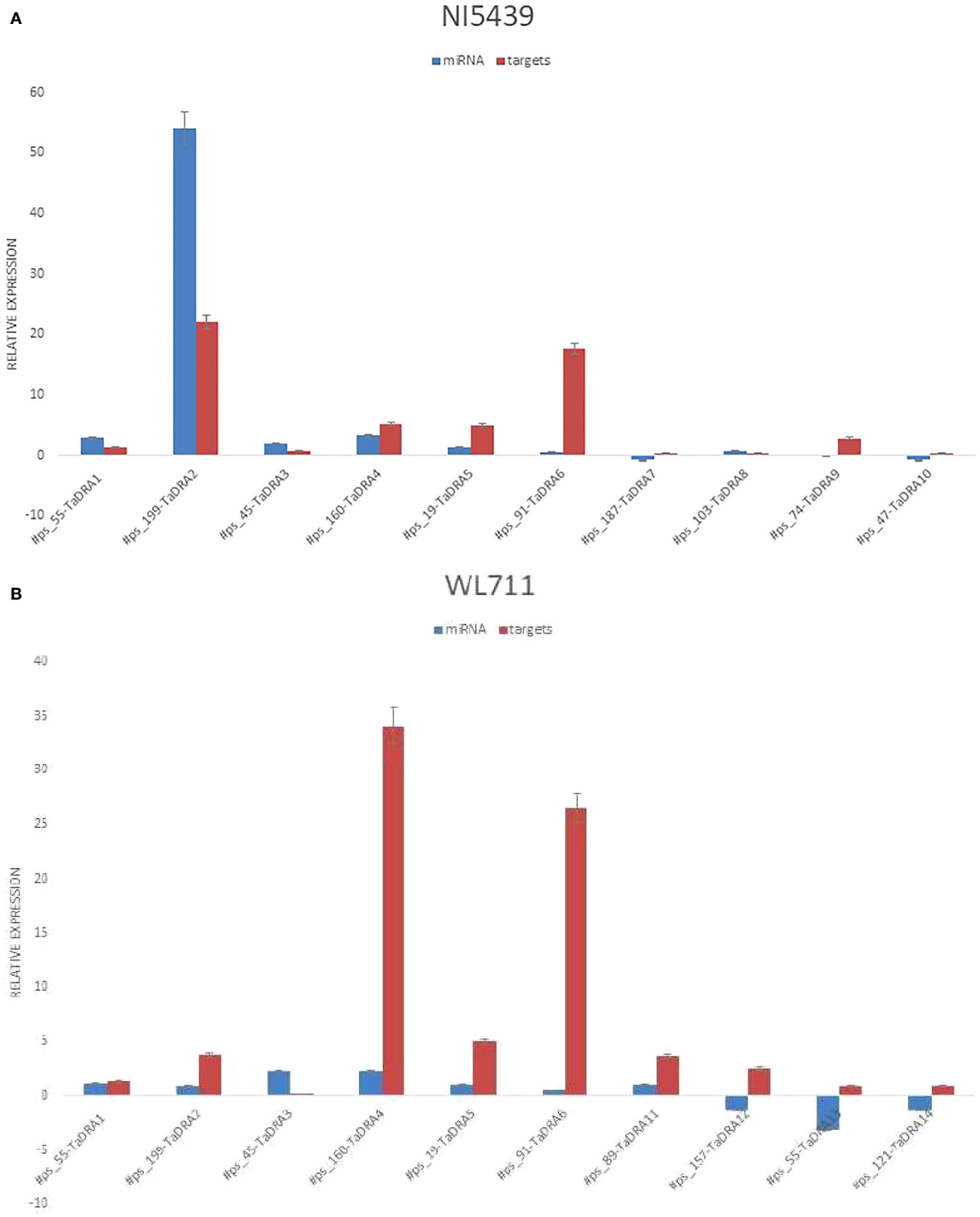


FIGURE 8
(A) Relative expression levels of miRNAs and their corresponding targets in root tissue of the drought-tolerant genotype NI5439. The orange colour represents the miRNA and blue represents the wheat genes. **(B)** Relative expression levels of miRNAs and their corresponding targets in the drought-susceptible genotype WL711 root tissue.

tae-miR395, tae-miR156, tae-miR398, and tae-miR319, were differentially expressed in response to drought stress. These findings align with prior research in wheat and other cereals, where these miRNAs have been implicated in drought tolerance mechanisms (Sunkar et al., 2007; Akdogan et al., 2016; Zhao et al., 2023a). The upregulation of tae-miR395 in this study is consistent with previous reports highlighting its role in sulfur metabolism and stress adaptation (Zhou et al., 2010; Wang et al., 2013). Similarly, tae-miR398, known to regulate Cu/Zn superoxide dismutase (CSD) genes involved in ROS detoxification, was significantly upregulated, reinforcing its importance in oxidative stress mitigation (Gupta et al., 2014). In addition to conserved miRNAs, 59 novel miRNAs were identified, with several exhibiting genotype-specific expression patterns. Notably, #ps_91 was highly expressed in the drought-tolerant genotype, suggesting their potential role in stress resilience. Similar genotype-dependent miRNA expression has been reported

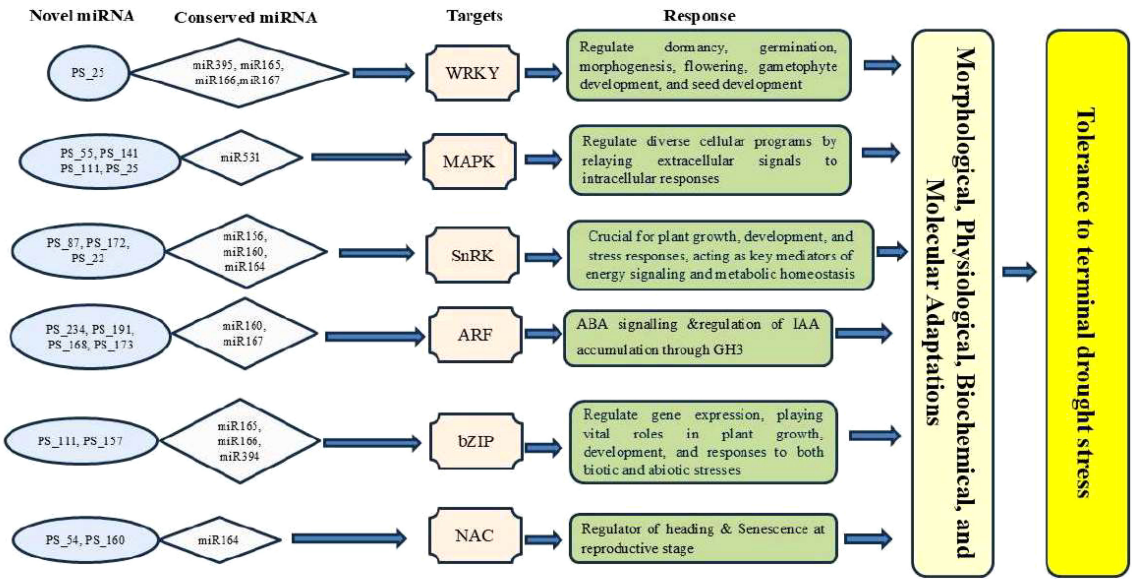


FIGURE 9
Illustrated depiction of the miRNA-mediated regulatory gene network involving known and novel miRNAs under drought stress at the reproductive stage in wheat.

in wheat, barley, and soybean, supporting the hypothesis that novel miRNAs contribute to adaptive stress responses (Gómez-Martín et al., 2023; Božić et al., 2024).

Hierarchical clustering and differential expression analysis revealed that several miRNAs exhibited significant upregulation or downregulation in response to drought stress. A total of 18 mature miRNAs were significantly expressed based on fold-change criteria. Among them, six miRNAs (tae-miR395a, tae-miR395b, tae-miR5049-3p, tae-miR5384-3p, tae-miR9664-3p, and tae-miR9666b-3p) were upregulated, whereas 12 miRNAs, including tae-miR156, tae-miR1135, tae-miR531, and tae-miR5200, were downregulated under drought conditions.The upregulation of tae-miR395 has been linked to enhanced sulfur metabolism and secondary metabolite biosynthesis, essential for stress adaptation (Kawakshima et al., 2009; Zhang J et al., 2022). The downregulation of tae-miR156 is of particular interest, as it has been previously associated with enhanced shoot growth and delayed flowering under stress conditions (Xu et al., 2021). These findings are consistent with prior studies in Arabidopsis and Oryza sativa, where miR156 negatively regulates SPL transcription factors, thereby influencing drought responses (Wang et al., 2011; Trindade et al., 2010). In contrast to previous studies where miR319 was downregulated under drought stress in wheat (Akdogan et al., 2016), this study observed its significant upregulation, particularly in tolerant genotypes. Given its role in targeting TCP transcription factors, this suggests that miR319 might be involved in modifying leaf morphology and cell wall biosynthesis to counteract stress effects.

The identified miRNA-target interactions highlight a complex regulatory network governing stress adaptation in wheat, aligning with previous findings in other plant species. The results demonstrated that several miRNAs target transcription factor (TF) families, including MYB, NAC, WRKY, and bZIP, which are critical regulators of gene expression under abiotic stress. Similar miRNA-mediated regulation of TFs has been reported in Arabidopsis and rice, where these TF families modulate drought-responsive pathways by controlling stress-inducible gene expression (Zhao et al., 2019; Li et al., 2021). The presence of hormone-related target genes, such as auxin response factors (ARFs) and abscisic acid (ABA)-responsive elements, further supports the role of miRNAs in integrating hormonal signaling with stress responses. Studies in maize and barley have demonstrated that miRNAs modulate ABA signaling to fine-tune stomatal regulation and osmotic balance under water-deficit conditions (Guan et al., 2020).

Gene ontology (GO) and KEGG pathway analyses revealed that the miRNA target genes are involved in critical biological processes, including transcriptional regulation, hormone signaling, ion transport, osmotic balance, and ROS detoxification.Several miRNAs were found to target transcription factors, including members of the WRKY, NAC, MYB, and bZIP families. These transcription factors are well-documented regulators of abiotic stress responses (He et al., 2016; Zhang Y et al., 2022). The downregulation of tae-miR169, which targets NF-YA transcription factors, suggests its role in modulating drought-responsive genes, consistent with previous findings in wheat and rice (Zhao et al., 2007; Bakhshi et al., 2014). The involvement of miRNAs in hormone signaling pathways was also evident. tae-miR159 was found to regulate MYB transcription factors involved in ABA signaling, highlighting its role in stress-adaptive hormone responses (Reyes and Chua, 2007; Pandey et al., 2013). Similarly, tae-miR160, which targets auxin response factors (ARFs), exhibited significant expression changes, suggesting its role in modulating root development and drought adaptation (Ding et al., 2013). One of the most striking findings was the regulation of ROS scavenging and oxidative stress pathways. The upregulation of tae-miR398, which regulates superoxide dismutase (SOD), aligns with previous

studies demonstrating miRNA-mediated control of oxidative stress responses in wheat and rice (Zhan and Meyers, 2023; Zhou et al., 2024). This supports the hypothesis that miRNA-regulated ROS detoxification plays a central role in drought tolerance.The identification of drought-responsive miRNAs provides a foundation for developing stress-resilient wheat varieties through molecular breeding and genetic engineering. The differential expression of miRNAs between drought-tolerant and drought-sensitive genotypes suggests that miRNAs could serve as biomarkers for selecting drought-adaptive traits in breeding programs.Several miRNAs identified in this study, particularly those regulating transcription factors, hormone signaling genes, and antioxidant enzymes, could be targeted for CRISPR/Cas9-based genome editing to enhance stress tolerance (Jaganathan et al., 2018; Tang and Chu, 2023). Based on our findings, we present a model illustrating miRNA-mediated gene regulation and stress responses in wheat under reproductive stage drought stress (Figure 9). Future studies should focus on functional validation of key miRNAs using transgenic approaches and miRNA knockout strategies to elucidate their precise roles in drought adaptation.Furthermore, integrating miRNA expression data with proteomics, metabolomics, and physiological measurements could provide a more comprehensive understanding of drought stress regulatory mechanisms in wheat. Multi-omics approaches will be crucial in unravelling the complex interplay between miRNAs and their targets, ultimately aiding in the development of climate-resilient crops (Zhao et al., 2023b).

5 Conclusion

In this study, four sRNA libraries were constructed from roots of drought-tolerant NI-5439 and drought-sensitive WL-711 wheat genotype from control and drought stress conditions at the booting stage. In total, 306 conserved and 58 novel microRNAs were identified from the four libraries. After computational expression analysis of mature miRNAs, 18 miRNAs showed significant changes in the expression after stress treatment. For the first time, 15 conserved miRNAs were emerged as drought-responsive in this study. The predicted targets of novel miRNAs were genes involved in gene silencing by RNA, DNA methylation, histone modification and chromatin modification. This study has significantly expanded the number of novel as well as drought-responsive miRNAs in wheat.

Data availability statement

The data have presented in the study are deposited in the NCBI, accession number is PRJNA1012115.

Author contributions

PS: Conceptualization, Data curation, Funding acquisition, Project administration, Supervision, Writing – original draft, Writing – review & editing. SM: Formal analysis, Investigation, Methodology, Software, Writing – original draft. AK: Validation, Writing – review & editing. OA: Writing – review & editing. RT: Funding acquisition, Writing – review & editing.

Funding

The author(s) declare that financial support was received forthe research and/or publication of this article. PS acknowledge the financial assistance from Indian Council of Agricultural research, New Delhi and ICAR-IIWBR Core projects to carry out 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.

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.2025.1581542/full#supplementary-material

References

Ahmad, P., Abd Allah, E. F., Hashem, A., Sarwat, M., and Gucel, S. (2023). Oxidative stress and antioxidant defense in wheat under drought conditions: Molecular and physiological perspectives. *Environ. Exp. Bot.* 208, 105181. doi: 10.1016/j.envexpbot.2023.105181

Akdogan, G., Tufekci, E. D., Uranbey, S., and Unver, T. (2016). miRNA-based drought regulation in wheat. *Funct. Integr. Genomics* 16, 221–233. doi: 10.1007/s10142-016-0476-7

Bailey, J., Oliveri, A., and Levin, E. (2013). A PCR-based method for detection and quantification of small RNAs. *Biochem. Biophys. Res. Commun.* 23, 1–7. doi: 10.1016/j.bbrc.2006.10.105

Bakhshi, B., Salekdeh, G. H., Bihamta, M. R., and Tohidfar, M. (2014). Characterization of three key MicroRNAs in rice root architecture under drought stress using in silico analysis and quantitative real-time PCR. *Biosci. Biotechnol. Res. Asia* 11, 555–565. doi: 10.13005/bbra/1306

Berahim, Z., Ismail, M. R., and Rahman, M. A. (2021). Physiological and biochemical responses of wheat to drought stress: Insights into ROS scavenging and photosynthetic efficiency. *J. Plant Physiol.* 263, 153468. doi: 10.1016/j.jplph.2021.153468

Božić, M., Ignjatović Micić, D., Delić, N., and Nikolić, A. (2024). Maize miRNAs and their putative target genes involved in chilling stress response in 5-day old seedlings. *BMC Genomics* 25, 1–17. doi: 10.1186/s12864-024-10403-1

Chen, H., Sun, W., and Zhang, L. (2023). Role of brassinosteroids in wheat drought resilience: Osmotic regulation and stress tolerance. *Plant Mol. Biol.* 112, 431–445. doi: 10.1007/s11103-023-01459-5

Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21 (18), 3674–3676. doi: 10.1093/bioinformatics/bti610

Ding, D., Zhang, L., Wang, H., Liu, Z., Zhang, Z., and Zheng, Y. (2013). Differential expression of miRNAs in response to salt stress in maize roots. *Ann. Bot.* 111, 767–779. doi: 10.1093/aob/mct043

Friedländer, M. R., MacKowiak, S. D., Li, N., Chen, W., and Rajewsky, N. (2012). MiRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res.* 40, 37–52. doi: 10.1093/nar/gkr688

Gómez-Martín, C., Zhou, H., Medina, J. M., Aparicio-Puerta, E., Shi, B., and Hackenberg, M. (2023). Genome-wide analysis of microRNA expression profile in roots and leaves of three wheat cultivars under water and drought conditions. *Biomolecules* 13 (3), 440. doi: 10.3390/biom13030440

Guan, Q., Wu, J., Zhang, Y., Jiang, C., Liu, R., Chai, C., et al. (2020). Crosstalk between microRNAs and ABA signaling pathways in plant drought resistance. *Cells* 9, 1757. doi: 10.3390/cells9081757

Gupta, O. P., Meena, N. L., Sharma, I., and Sharma, P. (2014). Differential regulation of microRNAs in response to osmotic, salt, and cold stresses in wheat. *Mol. Biol. Rep.* 41 (7), 4623–4629. doi: 10.1007/s11033-014-3363-1

He, H., Dong, Q., Shao, Y., Jiang, H., Zhu, S., Cheng, B., et al. (2016). Genome-wide survey and characterization of the WRKY gene family in *Populus trichocarpa*. *Plant Cell Rep.* 35, 21–37. doi: 10.1007/s00299-015-1866-3

Jaganathan, D., Ramasamy, K., Sellamuthu, G., Jayabalan, S., and Venkataraman, G. (2018). CRISPR for crop improvement: An update review. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.00985

Kawashima, C. G., Yoshimoto, N., Maruyama-Nakashita, A., Tsuchiya, Y. N., Saito, K., Takahashi, H., et al. (2009). Sulphur starvation induces the expression of microRNA-395 and one of its target genes but in different cell types. *The Plant J.* 57, 313–321. doi: 10.1111/j.1365-313X.2008.03690.x

Kaur, A., Gupta, O. P., Meena, N. L., Grewal, A., and Sharma, P. (2017). Comparative temporal expression analysis of micrornas and their target genes in contrasting wheat genotypes during osmotic stress. *Appl. Biochem. Biotechnol.* 181, 613–626. doi: 10.1007/s12010-016-2236-z

Kaur, R., Singh, J., and Dhillon, B. S. (2023a). Addressing climate challenges in wheat production: Advances in genetic and agronomic approaches. *Agron. J.* 115, 341–356. doi: 10.2134/agronj2022.05.0275

Kim, J. S., Kim, J., and Lee, H. (2020). Drought-induced oxidative stress and its impact on photosynthesis and growth in wheat: Mechanisms and mitigation strategies. *Plant Physiol. Biochem.* 150, 178–187. doi: 10.1016/j.plaphy.2020.03.005

Kozomara, A., Birgaoanu, M., and Griffiths-Jones, S. (2019). MiRBase: From microRNA sequences to function. *Nucleic Acids Res.* 47, D155–D162. doi: 10.1093/nar/gky1141

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10 (3), R25. doi: 10.1186/gb-2009-10-3-r25

Li, H., Liu, H., Pei, T., Zhang, Z., Yang, Y., and Liu, X. (2021). MicroRNA-mediated hormone signaling and transcription factor regulatory networks in plant drought adaptation. *Int. J. Mol. Sci.* 22, 4694. doi: 10.3390/ijms22094694

Liu, Y., Shi, A., Chen, Y., Xu, Z., and Yao, Y. (2023). Beneficial microorganisms: Regulating growth and defense for plant welfare. *Plant Biotechnol. J.* 22, 14554. doi: 10.1111/pbi.14554

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

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, 1–21. doi: 10.1186/s13059-014-0550-8

Lu, C., Meyers, B. C., and Green, P. J. (2007). Construction of small RNA cDNA libraries for deep sequencing. *Methods* 43, 110–117. doi: 10.1016/j.ymeth.2007.05.002

Meyers, B. C., Axtell, M. J., Bartel, B., Bartel, D. P., Baulcombe, D., Bowman, J. L., et al. (2008). Criteria for annotation of plant microRNAs. *Plant Cell* 20, 3186–3190. doi: 10.1105/tpc.108.064311

Mishra, P., Kumar, R., Singh, J., and Sharma, S. (2023a). Genome-wide expression studies reveal miRNA-mediated regulatory networks enhancing drought resilience in wheat. *J. Exp. Bot.* 74, 1956–1971. doi: 10.1093/jxb/erad065

Narayanan, S., Mohan, A., Gill, K. S., and Vara Prasad, P. V. (2014). Variability of root traits in spring wheat germplasm. *PloS One* 9 (6), e100317. doi: 10.1371/journal.pone.0100317

Nelson, G. C., Rosegrant, M. W., Koo, J., Robertson, R., Sulser, T., Zhu, T., et al. (2014). Climate change effects on agriculture: Economic impacts, adaptation, and mitigation. *Food Policy* 43, 18–28. doi: 10.1016/j.foodpol.2013.10.010

Pandey, R., Joshi, G., Bhardwaj, A. R., Agarwal, M., Katiyar-Agarwal, S., and Jagannath, A. (2013). A comprehensive genome-wide study on tissue-specific and abiotic stress-specific miRNAs in *Glycine max*. *Mol. Biol. Rep.* 40, 3101–3111. doi: 10.1007/s11033-012-2390-4

Parmar, S., Saharan, K., Kumar, P., and Singh, S. (2020). Climate change and its impact on wheat production: Strategies for mitigation and adaptation. *Environ. Sci. Pollution Res.* 27, 18192–18210. doi: 10.1007/s11356-020-08492-3

Ramachandran, S., Othman, S. M. I. S., Mustaffa, A. F., and Che-Othman, M. H. (2020). Overview of repressive miRNA regulation by short tandem target mimic (STTM): Applications and impact on plant biology. *Plants* 12, 669. doi: 10.3390/plants12030669

Reyes, J. L., and Chua, N. H. (2007). ABA induction of miR159 controls transcript levels of two MYB factors during Arabidopsis seed germination. *Plant J.* 49, 592–606. doi: 10.1111/j.1365-313X.2006.02980.x

Saroha, M., Arya, A., Singh, G., and Sharma, P. (2024). Genome-wide expression analysis of novel heat-responsive microRNAs and their targets in contrasting wheat genotypes at reproductive stage under terminal heat stress. *Front. Plant Sci.* 15. doi: 10.3389/fpls.2024.1328114

Sharma, P., Kumar, R., and Bhardwaj, R. (2023). MAPK signaling and hormonal crosstalk in drought stress tolerance of wheat: Mechanistic insights and applications. *Front. Plant Sci.* 14. doi: 10.3389/fpls.2023.1208759

Singh, V., van Oosterom, E. J., Jordan, D. R., Hunt, C. H., and Hammer, G. L. (2011). Genetic variability and control of nodal root angle in sorghum. *Crop Sci.* 51, 2011–2020. doi: 10.2135/cropsci2011.01.0038

Sun, J., Wang, P., and Li, X. (2023). Cytokinins and their role in mitigating drought-induced senescence in wheat: Molecular mechanisms and agricultural applications. *J. Plant Growth Regul.* 42, 1023–1038. doi: 10.1007/s00344-023-10782-6

Sunkar, R., Chinnusamy, V., Zhu, J., and Zhu, J. K. (2007). Small RNAs as big players in plant abiotic stress responses and nutrient deprivation. *Trends Plant Sci.* 12, 301–309. doi: 10.1016/j.tplants.2007.05.001

Sunkar, R., Kapoor, A., and Zhu, J. K. (2012a). Posttranscriptional induction of two Cu/Zn superoxide dismutase genes in Arabidopsis is mediated by downregulation of miR398 and important for oxidative stress tolerance. *Plant Cell* 18, 2051–2065. doi: 10.1105/tpc.106.041673

Sunkar, R., Li, Y. F., and Jagadeeswaran, G. (2012b). Functions of microRNAs in plant stress responses. *Trends Plant Sci.* 17, 196–203. doi: 10.1016/j.tplants.2012.01.010

Tang, M., and Chu, C. (2023). MicroRNAs in crop improvement: Fine-tuners for complex traits. *New Phytol.* 237, 1688–1703. doi: 10.1111/nph.18600

Tian, T., Liu, Y., Yan, H., You, Q., Yi, X., Du, Z., et al. (2017). AgriGO v2.0: a GO analysis toolkit for the agricultural community update. *Nucleic Acids Res.* 45 (W1), W122–W129. doi: 10.1093/nar/gkx382

Trindade, I., Capitão, C., Dalmay, T., Fevereiro, M. P., and Santos, D. M. (2010). miR398 and miR408 are up-regulated in response to water deficit in *Medicago truncatula*. *Plant Cell Environ.* 33, 1582–1597. doi: 10.1111/j.1365-3040.2010.02165.x

Wang, Y., Chen, J., and Zhang, Q. (2022). The interplay between abscisic acid and jasmonic acid in drought adaptation: Regulatory networks and physiological implications. *J. Exp. Bot.* 73, 2857–2870. doi: 10.1093/jxb/erac024

Wang, J. W., Czech, B., and Weigel, D. (2011). miR156-regulated SPL transcription factors define an endogenous flowering pathway in Arabidopsis thaliana. *Cell* 138, 738–749. doi: 10.1016/j.cell.2009.06.014

Wang, Y., Xu, L., Zhu, X., Zha, S., Yu, R., and Sun, L. (2013). Identification of microRNAs and their targets in wheat (Triticum aestivum L.) under phosphorus deprivation. *J. Plant Physiol.* 170 (18), 1668–1674. doi: 10.1016/j.jplph.2013.06.012

Wang, X., Zhang, H., Li, Y., and Zhao, C. (2023). Advances in breeding drought-tolerant wheat varieties: Molecular and physiological perspectives. *Plant Physiol. Biochem.* 201, 107611. doi: 10.1016/j.plaphy.2023.107611

Xu, M. Y., Zhang, L., Li, W. W., Hu, X. L., Wang, M. B., Fan, Y. L., et al. (2021). Stress-responsive miR156 improves drought resistance by modulating SPLs and metabolic processes in *Oryza sativa*. *J. Exp. Bot.* 72, 3754–3770. doi: 10.1093/jxb/erab017

Zadoks, J. C., and Board, E. (1999). Data sheet highlights close coupled pumps. *World Pumps* 1999, 9. doi: 10.1016/s0262-1762(99)80614-2

Zhan, J., and Meyers, B. C. (2023). Stress adaptation and innovations: Emerging roles of miRNAs in plant genomes. *Trends Plant Sci.* 28, 66–82. doi: 10.1016/j.tplants.2022.08.010

Zhang, Y., Wang, L., Li, Z., and Peng, Y. (2022). Genome-wide identification and characterization of the WRKY gene family in wheat (*Triticum aestivum* L.). *BMC Genomics* 23, 310. doi: 10.1186/s12864-022-08560-2

Zhang, J., Wang, Q., Wang, D., Du, Y., Zhang, G., and Yang, L. (2022). miR395 promotes sulfur metabolism and secondary metabolism to enhance stress adaptation in wheat. *Int. J. Mol. Sci.* 23, 2054. doi: 10.3390/ijms23042054

Zhao, B., Ge, L., Liang, R., Li, W., Ruan, K., Lin, H., et al. (2007). Members of miR-169 family are induced by high salinity and transiently inhibit the NF-YA transcription factor. *BMC Mol. Biol.* 8, 52. doi: 10.1186/1471-2199-8-52

Zhao, Y., Liu, H., Wang, X., and Li, Z. (2024). Impact of global warming on wheat productivity: Current trends and future projections. *Nat. Climate Change* 14, 45–57. doi: 10.1038/s41558-024-01764-9

Zhao, Y., Wang, Y., Xu, J., and Li, X. (2023a). The role of conserved miRNAs in drought stress responses in wheat and cereals: Regulatory mechanisms and adaptive significance. *Front. Plant Sci.* 14. doi: 10.3389/fpls.2023.1189472

Zhao, Y., Xu, Z., Mo, Q., Zou, C., Li, W., Xu, Y., et al. (2019). Combined small RNA and degradome sequencing reveals microRNA regulatory network with MYB, NAC, WRKY, and bZIP transcription factors in maize. *Sci. Rep.* 9, 16328. doi: 10.1038/s41598-019-52856-1

Zhao, Y., Zhang, Y., and Wang, M. (2023b). Multi-omics approaches reveal the regulatory network of microRNAs in wheat under drought stress. *Plant Sci.* 327, 111543. doi: 10.1016/j.plantsci.2023.111543

Zhou, L., Liu, Y., Liu, Z., Kong, D., Duan, M., and Luo, L. (2010). Genome-wide identification and analysis of drought-responsive microRNAs in Oryza sativa. *J. Exp. Bot.* 61, 4157–4168. doi: 10.1093/jxb/erq237

Zhou, M., Chen, H., Wei, D., Ma, H., and Lin, J. (2024). miR398 modulates plant responses to drought stress through regulation of SOD and other ROS-scavenging enzymes in rice. *Plant Sci.* 315, 111123. doi: 10.1016/j.plantsci.2023.111123

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 31, 3406–3415. doi: 10.1093/nar/gkg595

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 19 June 2025
DOI 10.3389/fpls.2025.1611784

OPEN ACCESS

EDITED BY
Gurjeet Singh,
Texas A and M University, United States

REVIEWED BY
Tanmaya Kumar Sahu,
Indian Institute of Agricultural Biotechnology (ICAR), India
Ali Babar,
University of Florida, United States

*CORRESPONDENCE
Niranjan Baisakh
nbaisakh@agcenter.lsu.edu

RECEIVED 14 April 2025
ACCEPTED 27 May 2025
PUBLISHED 19 June 2025

CITATION
Pradhan AK, Gandham P, Rajasekaran K and Baisakh N (2025) Predictive prioritization of genes significantly associated with biotic and abiotic stresses in maize using machine learning algorithms. *Front. Plant Sci.* 16:1611784. doi: 10.3389/fpls.2025.1611784

Predictive prioritization of genes significantly associated with biotic and abiotic stresses in maize using machine learning algorithms

Anjan Kumar Pradhan[1], Prasad Gandham[1], Kanniah Rajasekaran[2] and Niranjan Baisakh[1*]

[1]School of Plant, Environmental and Soil Sciences, Louisiana State University Agricultural Center, Baton Rouge, LA, United States, [2]Southern Regional Research Center, USDA-ARS, New Orleans, LA, United States

Both biotic and abiotic stresses pose serious threats to the growth and productivity of crop plants, including maize worldwide. Identifying genes and associated networks underlying stress resistance responses in maize is paramount. A meta-transcriptome approach was undertaken to interrogate 39,756 genes differentially expressed in response to biotic and abiotic stresses in maize were interrogated for prioritization through seven machine learning (ML) models, such as support vector machine (SVM), partial least squares discriminant analysis (PLSDA), k-nearest neighbors (KNN), gradient boosting machine (GBM), random forest (RF), naïve bayes (NB), and decision tree (DT) to predict top-most significant genes for stress conditions. Improved performances of the algorithms via feature selection from the raw gene features identified 235 unique genes as top candidate genes across all models for all stresses. Three genes such as *Zm00001eb176680*, *Zm00001eb176940*, and *Zm00001eb179190* expressed as *bZIP* transcription factor 68, glycine-rich cell wall structural protein 2, and aldehyde dehydrogenase 11 (*ALDH11*), respectively were commonly predicted as top-most candidates between abiotic stress and combined stresses and were identified from a weighted gene co-expression network as the hub genes in the brown module. However, only one gene *Zm00001eb038720* encoding RNA-binding protein AU-1/Ribonuclease E/G, predicted by the PLSDA algorithm, was found commonly expressed under both biotic and abiotic stress. Genes involved in hormone signaling and nucleotide binding were significantly differentially regulated under stress conditions. These genes had an abundance of antioxidant responsive elements and abscisic acid responsive elements in their promoter region, suggesting their role in stress response. The top-ranked genes predicted to be key players in multiple stress resistance in maize need to be functional validated to ascertain their roles and further utilization in developing stress-resistant maize varieties.

KEYWORDS

a(biotic) stress, artificial intelligence, gene expression, maize, RNA-Seq

Introduction

Plants are constantly subjected to various biotic and abiotic stresses that have negative effects on the growth, development, and productivity of economically important crops including maize (*Zea mays* L.) (Ramegowda and Senthil-Kumar, 2015). Maize is a main grain, forage, and energy crop as well as a genetic model plant (Farooqi et al., 2022). It is one of the most important cereal crops cultivated worldwide, mainly in Africa and South America (Kimotho et al., 2019). After wheat and rice, maize is the most frequent cereal food in Mexico (Nazari et al., 2023). United States remains the largest producer of maize. According to USDA, corn production for the 2024/25 marketing year is projected to be around 377.63 million metric tons, a 1% reduction than last year, which is attributed to extreme drought and heat during the 2024 crop year (https://www.fas.usda.gov/data/production/commodity/0440000).

Abiotic stresses such as drought, cold, submergence, salinity, waterlogging, heavy metal contamination, or nutrient deficiency can reduce crop yield by more than >50% (Mallikarjuna et al., 2020). Similarly, biotic stresses, caused by living organisms such as bacteria, viruses, fungi, or nematodes, negatively affect the productivity of maize by approximately 10% (Nazari et al., 2023). Plants undergo genome-wide transcriptome reprogramming in response to external stressors, which leads to induction and/or repression of genes associated with various mechanisms at whole plant, organellar, cellular, and molecular levels. Recent advances in various omics technologies have accelerated the identification of genes and biological processes controlling stress responses in plants. Analysis of transcriptomes has made it possible to identify genes overexpressed/repressed under specific stress (Sharma et al., 2013). However, in field conditions, plants are repeatedly exposed to multiple stresses simultaneously, which requires the plants to exercise efficient molecular mechanisms to recognize a host of signals to effectively respond to more than one stress (Sharma et al., 2013). Both biotic and abiotic stress factors and their various combinations in natural conditions elicit modified stress responses in plants. In addition to several genes, transcription factors (TFs) are known to be significantly involved in stress response in plants (Zhu, 2002). Many TFs from AP2/IREP, bZIP, MYB, NAC, and WRKY families have been found to improve stress resistance by regulating the expression of other stress-responsive genes in plants (Qin et al., 2011). Thus, the identification and characterization of key genes that are co-expressed in plants' response to both abiotic and biotic stresses would provide targets for genetic strategies to improve stress tolerance.

Abbreviations: ML, machine learning; AI, artificial intelligence; SVM, support vector machine; PLSDA, partial least squares discriminant analysis; KNN, k-nearest neighbors; GBM, gradient boosting machine; RF, random forest; NB, naïve bayes; DT, decision tree; BE, batch effect; DEGs, differentially expressed genes; VarImp, Variable importance; REF, recursive feature elimination; ROC, receiver operating characteristic; AUC, area under the curve; MCC, Mathews correlation coefficient; VIP, variable importance in projection; WGCNA, weighted gene co-expression network analysis; ME, module eigengene; GS, gene significance; CREs, cis-regulatory elements.

Over the years, several stress-regulated genes have been identified in maize using both microarray and RNA-seq approaches (Hayford et al., 2024). However, deciphering unique genes responding to specific or multiple stresses requires significant computational maneuvers. Meta-analysis is recognized as a reasonable yet statistically powerful approach where the results from multiple independent studies can be combined to eliminate the challenges due to variations between individual studies (Ramasamy et al., 2008; Keel and Lindholm-Perry, 2022). Meta-analysis has been used successfully on transcriptomic data of several crops including maize to identify potentially top candidate genes that are regulated in plants to cope with stress (Baisakh et al., 2023; Hayford et al., 2024; Wang et al., 2022; Nazari et al., 2024).

The multidimensionality of RNA-seq and microarray data owing to the high number of variables and genes with minimal sample size necessitates a gene selection technique to find the most informative, expressed genes and remove the redundancy in the original space (Mahendran et al., 2020). Machine Learning (ML), which is a subset of artificial intelligence, focuses on training the algorithms based on available datasets thus enabling models to learn to make decisions on their own from data without explicit programming. ML uses feature extraction and selection as a reduction technique for classification performance to make decisions on the top features (Kira and Rendell, 1992). In this study, we conducted a meta-analysis for integrating RNA-seq data across independent studies on biotic and abiotic stresses in maize to predictably identify the top-most significantly differentially expressed genes using ML tools. Furthermore, we compared the results from multiple ML models and integrated gene co-expression network analysis to identify the most useful stress-responsive genes.

Materials and methods

RNA-seq data collection and feature counts

All RNA-seq datasets related to both biotic and abiotic stresses in maize (Supplementary Table 1) were searched online using the publicly available sequence repository database i.e., NCBI Gene Sequence Read Archive (https://www.ncbi.nlm.nih.gov/sra). Raw sequence reads were subjected to quality control and filtering following the method described earlier (Bedre et al., 2015). Clean sequence reads were mapped against the B73 reference genome NAM 5.0 using HISAT2 (Kim et al., 2019) with default parameters. Mapped reads were then counted for genomic features such as genes, chromosomal locations, etc. using the FeatureCounts program (Liao et al., 2014).

Read counts preprocessing and merging

Gene expression data are often associated with high inconsistency due to noise and pieces attributed to the differences in sample numbers, labels, experimental conditions, etc. To correct these

biases caused by non-biological conditions, data were normalized using "normalize.quantiles" in R package *preprocessCore* version 1.56.0 following Schadt et al. (2001). However, normalization procedures do not adjust the sample data for batch effects (BE) when merging batches of data from multiple experiments that contain large batch-to-batch variation. Therefore, BE correction was performed using the '*ComBat*' function within the SVA package in R as described earlier (Zhang et al., 2022), which uses empirical Bayes method that estimates the LS parameters (mean and variance) for each gene and merges information from multiple genes with similar expression attributes in each batch.

Machine learning gene selection approaches

Gene expression data in the form of classified attributes with feature counts of several thousands of genes with expression patterns from multiple stress samples were analyzed to identify the most important genes, which required extraction and selection of features with discriminating ability. For ML, 80% of the data was used as a training set and the remaining 20% as the test/validation set. The classification of maize samples for the gene expression under control and (a)biotic stress was used to select the ML models that can best identify the top stress-responsive genes. We used seven ML algorithms (i.e., SVM, support vector machine; PLSDA, partial least squares discriminant analysis, KNN: K-nearest neighbors, GBM, gradient boosting machine; RF, random forest; NB, naïve bayes; and DT, decision tree) for the identification of most informative genes from the differentially expressed genes (DEGs). Variable importance (VarImp) evaluation functions were grouped with and without the model information. A model-based approach is more closely tied to its performance, which can integrate the correlation structure between the predictors (genes in this case) into the importance calculation. Each gene is assigned a separate variable importance for each class in classification models where all importance measures are scaled at minimum and maximum values of 0 and 100, respectively. The area under the receiver operating characteristic (ROC) curve (AUC) values for each gene was obtained using the "*filterVarImp*" function.

The SVM algorithm was used to identify top genes with R package *e1071* version 1.7–6 using the function of SVM-radial and default code: RFE (x = data, y = as.numeric (as.factor (group)), sizes = c (seq (2, 40, by = 2)), RFE-Control = rfeControl (functions = caretFuncs, method = "cv"), methods = "svmRadial"). Another model, Random Forest (version 4.6–14), was implemented using the *RF* algorithm with the following parameters: ntree = 100–500 and mtry = 1 – 8 (Kim et al., 2022). The relevance score and ranking of the genes in RF and SVM were determined following the recursive feature elimination (REF) method as per the program manual.

The PLS-DA method was implemented in R package *PLSDA* with PLS regression where Y is a set of binary response variables describing the categories (control or stress) of a categorical variable on X, where X is the gene expression matrix using the equation of Pérez-Enciso and Tenenhaus (2003) that used the algorithm of Wold et al. (1983) to allow for missing values. We identified the top genes using variable importance in projection (*VIP*) for each gene (Eriksson et al., 1999). GBM, an ensemble method (Hastie et al., 2009) is also used for regression and classification methods with reduced variance and bias in simple prediction models (Hastie et al., 2009). The *caret* package in R was used with the GBM function for selecting the top genes. The *caret* package was also used for KNN model with 10-fold cross-validation and default parameter *tuneLength* =10 to select the top-ranked genes for different stresses. Similarly, the DT model used the R package *Rpart* to select the top genes through *mtree* function. Another method, Naïve Bayes version 0.9.7 was adopted for the NB algorithm with 10-fold cross-validation for identifying the most important genes. The performance of each ML model was determined by classification metrics such as accuracy, precision, specificity, sensitivity (recall), F1-score, Mathews correlation coefficient (MCC), and ROC derived from the confusion matrix following the equations described in Sabanci et al. (2022). The confusion matrix was prepared with the maize gene under control labeled as 1 and stress as -1.

Gene co-expression network analysis

Weighted gene co-expression network analysis (WGCNA) is used in systems biology to make clusters (modules) of highly correlated genes based on the module eigengene (ME) and to identify an intramodular hub gene. The expression values of genes after normalization and batch effect correction were used in the R package *WGCNA* version 1.66 (Langfelder and Horvath, 2008) for weighted co-expression network construction where the similar matrix between each pair of genes across all samples was evaluated based on the Pearson's correlation values. Modules were identified by the *blockwiseModules* function of the *WGCNA* package with default parameters and a tree cut height of 0.4. The modules were defined at a cut height of 0.98 and a size at 30. Similar modules were merged when dissimilarity of module eigengenes was <0.25. The function *signedKME* was used to calculate module eigengene values (KME) based on the correlation of the module eigengene (ME) with the corresponding gene. Modules with correlation value (r) >0.8 between genes and P-value <0.01 were considered as significant modules (de Silva et al., 2022). Gene significance was calculated based on the *p* value of the linear regression between the gene expression profile with multiple stress conditions. The hub genes in a module were identified based on gene significance value (GS) >0.5, module eigengene (KME) >0.8, and with maximum connections with other genes (Baisakh et al., 2023). Gene co-expression network was visualized using Cytoscape (version 3.10.3) software in R (Shannon et al., 2003; Baisakh et al., 2023).

Gene ontology and promoter analysis

The ontology of the genes was assigned by the singular enrichment analysis within AgriGO version 2 (https://

TABLE 1A Confusion matrices of machine learning algorithms used with gene expression values under biotic stress in maize.

Models	Accuracy	Specificity	Sensitivity	Precision	F1-score	MCC	FP Rate	ROC Area	PRC Area
NB	57	0.418	0.837	0.45	0.585	0.318	0.581	1.44	0.537
KNN	63.7	0.75	0.601	0.885	0.716	0.299	0.250	2.40	1.47
DT	52.5	0.459	0.581	0.565	0.573	0.034	0.540	1.07	0.972
GBM	71.8	0.687	0.746	0.726	0.736	0.434	0.528	1.41	0.973
SVM	76.6	0.809	0.729	0.813	0.555	0.538	0.190	3.83	1.11
RF	73	0.6	0.8	0.8	0.8	0.4	0.400	2.00	1.00
PLSDA	54.9	0.431	0.602	0.702	0.648	0.212	0.568	1.05	1.16

TABLE 1B Performance matrices of top genes predicted under biotic stress in maize.

Naive Bayes		GBM		DT		RF		SVM		K-NN		PLSDA	
Genes	AUC	Genes	rel.inf	Genes	rel.inf	Genes	AUC	Genes	RMSE	Genes	AUC	Genes	VIP-score
Zm00001eb034620	100	Zm00001eb024180	2.749	Zm00001eb110400	15.767	Zm00001eb088840	100	Zm00001eb019130	13.103	Zm00001eb034620	100	Zm00001eb176300	3.597
Zm00001eb019590	95.14	Zm00001eb034330	1.932	Zm00001eb087410	13.015	Zm00001eb034620	97.71	Zm00001eb034390	10.995	Zm00001eb145370	99.23	Zm00001eb021020	3.484
Zm00001eb136760	94.34	Zm00001eb164530	1.601	Zm00001eb097390	12.867	Zm00001eb158600	97.7	Zm00001eb090610	8.963	Zm00001eb034390	99.09	Zm00001eb047030	3.404
Zm00001eb071870	90.04	Zm00001eb071760	1.547	Zm00001eb034390	12.142	Zm00001eb090610	95.54	Zm00001eb181690	3.937	Zm00001eb070640	94.85	Zm00001eb038720	3.316
Zm00001eb145370	88.4	Zm00001eb068730	1.537	Zm00001eb088150	9.324	Zm00001eb034390	71.42	Zm00001eb088150	3.567	Zm00001eb037820	94.61	Zm00001eb168410	3.269
Zm00001eb150630	87.8	Zm00001eb029780	1.48	Zm00001eb160330	8.25	Zm00001eb042770	57.97	Zm00001eb115060	3.486	Zm00001eb166710	91.05	Zm00001eb137800	3.267
Zm00001eb067180	87.11	Zm00001eb084520	1.434	Zm00001eb019620	8.097	Zm00001eb151510	55.2	Zm00001eb110450	3.467	Zm00001eb063720	90.73	Zm00001eb100490	3.203
Zm00001eb103210	86.27	Zm00001eb046720	1.405	Zm00001eb171130	7.986	Zm00001eb010720	54.44	Zm00001eb184360	2.821	Zm00001eb090610	90.54	Zm00001eb096510	3.193
Zm00001eb050760	85.22	Zm00001eb027600	1.38	Zm00001eb034620	7.939	Zm00001eb078220	53.74	Zm00001eb100260	2.778	Zm00001eb150630	86.82	Zm00001eb022830	3.182
Zm00001eb157100	85.19	Zm00001eb103260	1.373	Zm00001eb053690	7.351	Zm00001eb002660	51	Zm00001eb109670	2.526	Zm00001eb028560	86.76	Zm00001eb143050	3.149
Zm00001eb143200	84.91	Zm00001eb075780	1.365	Zm00001eb034330	7.005	Zm00001eb144000	48.32	Zm00001eb146780	2.301	Zm00001eb067440	86.22	Zm00001eb056920	3.135
Zm00001eb097950	84.9	Zm00001eb156230	1.321	Zm00001eb034490	7.005	Zm00001eb088150	45.85	Zm00001eb167510	2.281	Zm00001eb043810	85.91	Zm00001eb111400	3.134
Zm00001eb156230	84.7	Zm00001eb061490	1.307	Zm00001eb043550	6.274	Zm00001eb143200	44.58	Zm00001eb167230	2.031	Zm00001eb020250	85.57	Zm00001eb065380	3.13
Zm00001eb088150	84.59	Zm00001eb155370	1.263	Zm00001eb015830	5.476	Zm00001eb063710	42.95	Zm00001eb097390	2.02	Zm00001eb071760	84.86	Zm00001eb016660	3.124

(Continued)

10.3389/fpls.2025.1611784

10.3389/fpls.2025.1611784

TABLE 1B Continued

Naive Bayes		GBM		DT		RF		SVM		K-NN		PLSDA	
Genes	AUC	Genes	rel.inf	Genes	rel.inf	Genes	AUC	Genes	RMSE	Genes	AUC	Genes	VIP-score
Zm00001eb002660	83.45	Zm00001eb133930	1.208	Zm00001eb016210	5.219	Zm00001eb094950	42.68	Zm00001eb020250	1.927	Zm00001eb099610	84.67	Zm00001eb081510	3.097
Zm00001eb110740	83.4	Zm00001eb003880	1.19	Zm00001eb018440	5.085	Zm00001eb019130	41.4	Zm00001eb088840	1.879	Zm00001eb075030	84.32	Zm00001eb098910	3.09
Zm00001eb182980	83.32	Zm00001eb083490	1.1849	Zm00001eb116160	5.085	Zm00001eb011650	39.35	Zm00001eb087410	1.842	Zm00001eb150350	84.28	Zm00001eb131030	3.089
Zm00001eb070860	83.29	Zm00001eb154470	1.182	Zm00001eb018350	4.87	Zm00001eb174340	39.35	Zm00001eb118770	1.74	Zm00001eb165610	83.8	Zm00001eb039370	3.087
Zm00001eb109860	83.07	Zm00001eb157410	1.175	Zm00001eb063710	4.662	Zm00001eb097390	37.74	Zm00001eb042770	1.587	Zm00001eb075420	83.65	Zm00001eb104430	3.087
Zm00001eb078870	82.68	Zm00001eb045900	1.166	Zm00001eb075030	4.662	Zm00001eb071400	37.58	Zm00001eb170900	1.555	Zm00001eb094950	83.55	Zm00001eb043420	3.085

MCC = Mathews correlation coefficient, ROC = receiver operating characteristic, PRC = precision recall curve.

systemsbiology.cau.edu.cn/agriGOv2/) and gene ontology enrichment was performed using Fisher's *t-test* (P<0.05) and FDR correction by the Hochberg method. Metabolic pathway enrichment analysis was performed using the DAVID tool version 6.7 (https://davidbioinformatics.nih.gov/tools.jsp). The top genes common between two or more models were used to extract 2000 bp upstream flanking region of sequence using the Ensembl Plants database (http://plants.ensembl.org). Promoter prediction and identification of cis-regulatory elements (CREs) within a promoter were performed using PlantCARE (https://bioinformatics.psb.ugent.be/webtools/plantcare/html/).

Results and discussion

DEGs under biotic and abiotic stress

Raw sequence reads of a total of 3,052 samples, which included 976 from biotic (bacteria, fungus, insect, nematode, and weed) and 2,076 from abiotic (drought, heat, cold, waterlogging, salt, nutrient, and mechanical wounding) stresses, were obtained from 52 RNA-seq studies conducted with 12 stress conditions of which five were biotic and seven were abiotic (Supplementary Table 1). Finally, 39,756 differentially expressed genes from 1,452 samples (451 from biotic and 1001 from abiotic) and 22 of 52 studies after normalization and batch effect correction, respectively were used in ML models sets to identify the topmost significant genes.

Identification of top stress-responsive genes by ML models

Identifying top stress-responsive DEGs was conducted for biotic or abiotic stress individually and in combination to identify top genes unique to a specific stress category and common between the stress categories.

Prediction of top genes under biotic stress

Based on the confusion matrix (Supplementary Table 2) of the seven ML models, SVM performed the best with the highest average accuracy of 76.6% followed by RF (73.0%) and GBM (71.8%) whereas NB, KNN, DT, and PLSDA performed very poorly with an average accuracy of 57% (Table 1A). SVM also had the highest specificity (0.81) and MCC (0.54) with precision (0.81) behind KNN (0.84) and sensitivity (0.73) behind NB (0.83), RF (0.80), and GBM (0.75). Interestingly, SVM had the lowest F1-score (0.56). However, MCC, which considers all four parameters in the confusion matrix, is considered a better performance matrix as compared to F1-score, which only considers precision/recall. DT had the worst performance in terms of accuracy (53.5%), sensitivity (0.58), and MCC (0.03), although it had a slightly better specificity over PLSDA and NB, higher precision than NB, and higher F1-score than SVM. DT was the least sensitive algorithm correctly detecting only 0.58 for positive samples.

TABLE 2A Confusion matrices of machine learning algorithms used with gene expression values under abiotic stress in maize.

Models	Accuracy (%)	Specificity	Sensitivity	Precision	F1-score	MCC	FP Rate	ROC Area	PRC Area
NB	80.30	0.893	0.675	0.813	0.748	0.589	0.107	6.30	1.20
KNN	80.00	0.767	0.932	0.495	0.647	0.580	0.397	2.34	0.531
DT	77.33	0.801	0.721	0.657	0.688	0.511	0.198	3.64	0.911
GBM	88.60	0.895	0.868	0.803	0.834	0.600	0.104	8.34	0.925
SVM	87.50	0.939	0.750	0.864	0.803	0.716	0.058	12.93	1.15
RF	87.00	0.897	0.840	0.790	0.984	0.725	0.111	7.56	0.940
PLSDA	78.10	0.853	0.644	0.697	0.695	0.507	0.146	4.41	1.08

TABLE 2B Performance matrices of top genes predicted under abiotic stress in maize.

Naive Bayes		GBM		DT		RF		SVM		K-NN		PLSDA	
Genes	AUC	Genes	rel.inf	Genes	rel.inf	Genes	AUC	Genes	RMSE	Genes	AUC	Genes	VIP-score
Zm00001eb012040	100	Zm00001eb146690	5.456	Zm00001eb012040	134.264	Zm00001eb112930	100	Zm00001eb012040	164.770	Zm00001eb160470	100	Zm00001eb133500	4.406
Zm00001eb146690	98.30	Zm00001eb058820	5.212	Zm00001eb112930	96.502	Zm00001eb021010	88.00	Zm00001eb160470	72.756	Zm00001eb176940	99.94	Zm00001eb027100	4.332
Zm00001eb176680	97.50	Zm00001eb012040	4.980	Zm00001eb160470	94.104	Zm00001eb012040	83.00	Zm00001eb176680	52.022	Zm00001eb012040	99.62	Zm00001eb130760	4.200
Zm00001eb021010	97.50	Zm00001eb044020	4.689	Zm00001eb098220	91.107	Zm00001eb176680	80.30	Zm00001eb201180	33.061	Zm00001eb126900	99.05	Zm00001eb117220	4.158
Zm00001eb160470	96.80	Zm00001eb176680	4.590	Zm00001eb050500	88.710	Zm00001eb146690	78.30	Zm00001eb146690	20.824	Zm00001eb050500	98.94	Zm00001eb099850	3.991
Zm00001eb044020	96.40	Zm00001eb176940	2.674	Zm00001eb179190	86.312	Zm00001eb058820	76.40	Zm00001eb248930	20.611	Zm00001eb044020	98.60	Zm00001eb104530	3.899
Zm00001eb112930	95.50	Zm00001eb021010	2.138	Zm00001eb151430	18.992	Zm00001eb160470	74.10	Zm00001eb192710	19.716	Zm00001eb176680	97.59	Zm00001eb068620	3.884
Zm00001eb050500	95.50	Zm00001eb075250	1.858	Zm00001eb072870	16.477	Zm00001eb044020	66.90	Zm00001eb238010	16.154	Zm00001eb179190	97.43	Zm00001eb090990	3.824
Zm00001eb176940	95.20	Zm00001eb030400	1.591	Zm00001eb065100	13.599	Zm00001eb098220	65.60	Zm00001eb203690	15.104	Zm00001eb146690	96.90	Zm00001eb115290	3.776
Zm00001eb151430	95.00	Zm00001eb151430	1.478	Zm00001eb156020	10.218	Zm00001eb176940	55.00	Zm00001eb050500	14.803	Zm00001eb151430	95.77	Zm00001eb149960	3.769
Zm00001eb058820	94.50	Zm00001eb156510	1.329	Zm00001eb171000	8.514	Zm00001eb018700	53.50	Zm00001eb176940	14.077	Zm00001eb041030	95.20	Zm00001eb037980	3.763
Zm00001eb179190	92.90	Zm00001eb119820	1.320	Zm00001eb143640	7.682	Zm00001eb076550	51.80	Zm00001eb021010	10.680	Zm00001eb076550	95.17	Zm00001eb046590	3.754
Zm00001eb126900	92.10	Zm00001eb050500	1.267	Zm00001eb076250	7.622	Zm00001eb151430	50.10	Zm00001eb250120	8.224	Zm00001eb058820	95.05	Zm00001eb000340	3.747
Zm00001eb041030	91.10	Zm00001eb074930	0.945	Zm00001eb019280	7.414	Zm00001eb179190	49.80	Zm00001eb017550	7.669	Zm00001eb112930	94.89	Zm00001eb114240	3.745

(Continued)

10.3389/fpls.2025.1611784

10.3389/fpls.2025.1611784

TABLE 2B Continued

Naive Bayes		GBM		DT		RF		SVM		K-NN		PLSDA	
Genes	AUC	Genes	rel.inf	Genes	rel.inf	Genes	AUC	Genes	RMSE	Genes	AUC	Genes	VIP-score
Zm00001eb098220	90.50	Zm00001eb057510	0.939	Zm00001eb136970	6.664	Zm00001eb078640	47.60	Zm00001eb112930	7.260	Zm0000leb021010	94.61	Zm00001eb038720	3.726
Zm00001eb018700	90.10	Zm00001eb171000	0.903	Zm00001eb154950	6.590	Zm00001eb050500	45.50	Zm00001eb189080	5.677	Zm0000leb005840	94.24	Zm00001eb066950	3.687
Zm00001eb117180	89.80	Zm00001eb111020	0.889	Zm00001eb013080	6.219	Zm00001eb148130	38.10	Zm00001eb044020	5.583	Zm0000leb124290	92.63	Zm00001eb148300	3.677
Zm00001eb005840	89.80	Zm00001eb015730	0.873	Zm00001eb074360	6.219	Zm00001eb040280	37.10	Zm00001eb151430	4.550	Zm0000leb018700	92.50	Zm00001eb026650	3.673
Zm00001eb076550	89.80	Zm00001eb080960	0.858	Zm00001eb127040	6.219	Zm00001eb117180	36.90	Zm00001eb043060	4.516	Zm0000leb013780	91.83	Zm00001eb124020	3.671
Zm00001eb124290	89.70	Zm00001eb154060	0.816	Zm00001eb128310	6.004	Zm00001eb179280	35.00	Zm00001eb243850	4.176	Zm0000leb075250	91.53	Zm00001eb128720	3.660

MCC, Mathews correlation coefficient; ROC, receiver operating characteristic; PRC, precision recall curve.

The top 20 genes predicted by each model resulted in a total of 111 unique top significantly differentially regulated genes by all seven models (Table 1B). Of these, 16 genes were predicted by at least two models. Four genes, *Zm00001eb034390*, *Zm00001eb088150*, *Zm00001eb042770*, and *Zm00001eb097390* predicted as top genes by two or more models including the two high-performing models, SVM and RF, were considered the most significantly differentially expressed genes under biotic stress.

Prediction of top genes under abiotic stress

The confusion matrix (Supplementary Table 2) for the abiotic stress responsive genes revealed that the highest average accuracy of 88.6% was obtained by GBM, closely followed by SVM (87.5%) and RF (87.0%) algorithms (Table 2A). RF had the highest MCC value (0.72) followed by SVM at 0.71 and GBM (0.6). SVM had the highest specificity (0.94) followed by RF at 0.89 and precision (0.86). Altogether, 68 unique genes were identified as most informative by all seven models. Twenty-one genes were consistently predicted as top abiotic stress-related genes by at least two models of which 12 were common in four or more models, which included GBM, SVM, RF and KNN models that had higher accuracy and/or model performance matrices compared to other models (Table 2B). Interestingly, only one gene *Zm00001eb038720*, predicted by PLSDA, was consistent between the biotic and abiotic stress conditions.

Prediction of top genes under combined stress conditions

When the genes responsive to biotic and/or abiotic stress conditions based on the datasets in the literature were combinedly used for prediction by seven models, RF outperformed others with the highest accuracy (81.7%) and other performance matrices except for sensitivity (0.75), which was behind SVM (0.80) and GBM (0.77) (Table 3A; Supplementary Table 2). SVM predicted the top 20 genes with the highest sensitivity, an accuracy of 81.0%, F1-score (0.77) and MCC (0.61) second to only RF. GBM also performed good with 79.0% accuracy, nearly equal specificity and same sensitivity as SVM. NB was the worst predictor algorithm with the lowest values recorded for accuracy as well as other parameters. A total of 83 unique top significant genes were reported for combined stress by the seven models of which 11 genes were commonly predicted by four or more models. Among biotic and combined stress conditions, 23 genes were found common. On the other hand, only two genes were found within biotic and combined stress conditions (Table 3B).

Taken together, SVM, RF, and GBM were identified as the best models in predicting top significant (a)biotic stress responsive genes with high accuracy in our study. However, Nazari et al. (2023) found KNN (82.0%) and Ensemble (85.7%) to be more accurate for predicting biotic stress tolerance genes while modeling gene expression data from microarray studies in maize. Interestingly, none of the significant genes identified by these authors matched the top genes selected by the seven models used in our study.

TABLE 3A Confusion matrices of machine learning algorithms used with gene expression values under combined stress conditions in maize.

Models	Accuracy (%)	Specificity	Sensitivity	Precision	F1-score	MCC	FP Rate	ROC Area	PRC Area
NB	65.10	0.680	0.591	0.466	0.523	0.232	0.319	1.85	0.788
KNN	72.00	0.725	0.711	0.537	0.612	0.401	0.275	2.58	0.755
DT	73.40	0.779	0.672	0.689	0.658	0.453	0.220	3.05	1.02
GBM	79.00	0.814	0.772	0.753	0.763	0.584	0.185	4.17	0.975
SVM	81.00	0.815	0.803	0.746	0.773	0.612	0.184	4.36	0.929
RF	81.70	0.865	0.755	0.813	0.783	0.627	0.134	5.63	1.07
PLSDA	78.00	0.840	0.690	0.760	0.720	0.549	0.158	4.36	1.10

TABLE 3B Performance matrices of top genes predicted under combined stress conditions in maize.

Naive Bayes		GBM		DT		RF		SVM		K-NN		PLSDA	
Genes	AUC	Genes	rel.inf	Genes	rel.inf	Genes	AUC	Genes	RMSE	Genes	AUC	Genes	VIP-score
Zm00001eb050500	100	Zm00001eb112930	3.789	Zm00001eb176940	142.216	Zm00001eb076550	100	Zm00001eb050500	2.783	Zm00001eb050500	100	Zm00001eb130760	1.935
Zm00001eb146690	99.32	Zm00001eb179190	3.317	Zm00001eb018700	108.111	Zm00001eb179190	99.66	Zm00001eb176680	2.383	Zm00001eb176940	95.78	Zm00001eb163120	1.931
Zm00001eb176940	98.20	Zm00001eb050500	3.109	Zm00001eb041030	105.736	Zm00001eb050500	96.40	Zm00001eb176940	2.372	Zm00001eb146690	94.84	Zm00001eb162400	1.926
Zm00001eb112930	96.84	Zm00001eb126900	3.048	Zm00001eb126900	105.736	Zm00001eb160470	95.19	Zm00001eb126900	2.240	Zm00001eb160470	93.99	Zm00001eb104530	1.925
Zm00001eb058820	94.47	Zm00001eb176680	2.789	Zm00001eb050500	103.741	Zm00001eb176940	92.86	Zm00001eb041030	2.228	Zm00001eb112930	92.88	Zm00001eb113530	1.924
Zm00001eb176680	93.17	Zm00001eb176940	2.208	Zm00001eb075230	99.466	Zm00001eb126900	85.99	Zm00001eb018700	2.224	Zm00001eb176680	92.58	Zm00001eb177670	1.922
Zm00001eb160470	92.55	Zm00001eb146690	1.650	Zm00001eb012040	25.994	Zm00001eb112930	80.01	Zm00001eb001220	1.902	Zm00001eb058820	89.88	Zm00001eb165700	1.921
Zm00001eb179190	90.05	Zm00001eb058820	1.326	Zm00001eb071400	21.751	Zm00001eb176680	73.21	Zm00001eb003440	1.900	Zm00001eb179190	88.72	Zm00001eb149960	1.921
Zm00001eb033200	87.67	Zm00001eb154060	1.006	Zm00001eb005840	21.107	Zm00001eb018700	73.05	Zm00001eb050470	1.870	Zm00001eb154060	88.22	Zm00001eb050770	1.920
Zm00001eb154060	87.42	Zm00001eb160470	0.982	Zm00001eb116880	20.899	Zm00001eb146690	65.41	Zm00001eb160470	1.667	Zm00001eb041030	86.33	Zm00001eb028190	1.918
Zm00001eb041030	87.23	Zm00001eb021720	0.841	Zm00001eb021010	20.275	Zm00001eb037690	63.90	Zm00001eb058820	1.649	Zm00001eb156130	84.24	Zm00001eb028490	1.918
Zm00001eb018180	85.09	Zm00001eb054980	0.833	Zm00001eb031210	20.275	Zm00001eb058820	62.44	Zm00001eb146690	1.569	Zm00001eb146500	81.84	Zm00001eb089070	1.918
Zm00001eb156130	84.54	Zm00001eb042770	0.828	Zm00001eb155430	20.067	Zm00001eb012040	60.49	Zm00001eb093590	1.451	Zm00001eb038010	81.69	Zm00001eb138650	1.917
Zm00001eb138530	83.53	Zm00001eb140230	0.772	Zm00001eb166600	17.528	Zm00001eb154060	59.43	Zm00001eb004630	1.443	Zm00001eb033200	81.28	Zm00001eb154420	1.917

(Continued)

10.3389/fpls.2025.1611784

10.3389/fpls.2025.1611784

TABLE 3B Continued

Naive Bayes		GBM		DT		RF		SVM		K-NN		PLSDA	
Genes	AUC	Genes	rel.inf	Genes	rel.inf	Genes	AUC	Genes	RMSE	Genes	AUC	Genes	VIP-score
Zm00001eb146500	83.38	Zm00001eb079740	0.714	Zm00001eb019280	16.114	Zm00001eb075230	58.23	Zm00001eb040610	1.434	Zm00001eb138530	81.13	Zm00001eb042970	1.917
Zm00001eb003440	83.01	Zm00001eb057750	0.686	Zm00001eb171940	14.569	Zm00001eb041030	46.05	Zm00001eb129480	1.410	Zm00001eb016200	80.71	Zm00001eb124300	1.917
Zm00001eb004630	82.89	Zm00001eb020980	0.671	Zm00001eb171930	14.341	Zm00001eb148130	45.43	Zm00001eb111990	1.398	Zm00001eb040610	80.69	Zm00001eb172930	1.916
Zm00001eb010960	82.43	Zm00001eb146390	0.670	Zm00001eb166670	14.114	Zm00001eb179280	42.10	Zm00001eb098220	1.363	Zm00001eb179280	80.67	Zm00001eb033410	1.916
Zm00001eb014630	82.43	Zm00001eb016200	0.649	Zm00001eb166610	13.886	Zm00001eb020260	41.89	Zm00001eb101050	1.354	Zm00001eb078640	80.29	Zm00001eb042710	1.916
Zm00001eb163520	82.41	Zm00001eb102720	0.642	Zm00001eb170070	13.886	Zm00001eb050850	38.64	Zm00001eb112930	1.349	Zm00001eb014630	80.08	Zm00001eb131600	1.916

MCC, Mathews correlation coefficient; ROC, receiver operating characteristic; PRC, precision recall curve.

Significant modules and potential hub genes

The weighted gene co-expression network analysis based on a height cut-off at 0.25 to merge the modules, detected four modules of which Turquoise and Brown modules with 438 and 14 genes, respectively were significant with r >0.8 and GS (P-value¾0.01) (Supplementary Table 3; Figure 1a). Three significant hub genes, *Zm00001eb176680, Zm00001eb176940*, and *Zm00001eb179190* with KME>0.8, were identified in the brown module (Figure 1b) for abiotic stress and one gene *Zm00001eb150630* (turquoise model) for biotic stress (Figure 1c) that were predicted by two or more models.

Functional involvement of the top highly significant genes

GO analysis of top genes representing conserved up and down-regulation under stress conditions identified 47 and 3 significant GO terms (p-value<0.001; FDR <0.05) for abiotic and biotic stress, respectively whereas 17 significant GO terms were identified for combined stress (Supplementary Table 4). The most significantly enriched molecular function with the highest number of genes was associated with nucleotide binding for biotic stress whereas binding followed by response to stimulus were the most enriched processes under abiotic stress. On the other hand, cell communication, cellular response to alcohol, lipid, external stimuli, and abscisic acid, and hormone signaling were the two most enriched biological processes under combined stress conditions. The biological pathways involving the top-most significant genes involved are presented in Supplementary Table 5. In biotic stress, genes involved in calcium ion binding were the most enriched whereas genes in abscisic acid activated signaling pathway, cation binding and myosin phosphate activity were the most significant in abiotic stress. Genes associated with seed nutrient storage activity, cation binding, and plant hormone signal transduction were significant under combined, biotic and abiotic stresses.

Of the four top-most significant genes associated with biotic stress response and predicted by the two high-performing models SVM and RF, *Zm00001eb034390* (GO:0005509) coding for EF-hand 1 calcium binding protein (1.23-fold), *Zm00001eb088150* (GO:0098754) for UDP-glucuronosyl/UDP-glucosyltransferase (0.74-fold), *Zm00001eb042770* (GO:0005634) for ribonuclease 3-like protein 3 (0.23) were significantly upregulated, whereas gene *Zm00001eb097390* (GO:0006457) encoding GrpE nucleotide exchange factor was downregulated (-0.14) under fungal infection (Supplementary Table 6) (Shu et al., 2017; Kebede et al., 2018; Shi et al., 2018; Han et al., 2020; Lambarey et al., 2020; Musungu et al., 2020; He et al., 2021; Liu et al., 2021; Schurack et al., 2021). *Zm00001eb088150* has been shown to be involved in the detoxification of several exogenous and endogenous compounds, and it plays multiple roles in plant responses to biotic as well as abiotic stresses (Gharabli et al., 2023) providing protection against mycotoxins (Wetterhorn, 2018), pathogens, drought, heat, cold, and salinity (Van Aken, 2008; Meale et al., 2015). UDP-

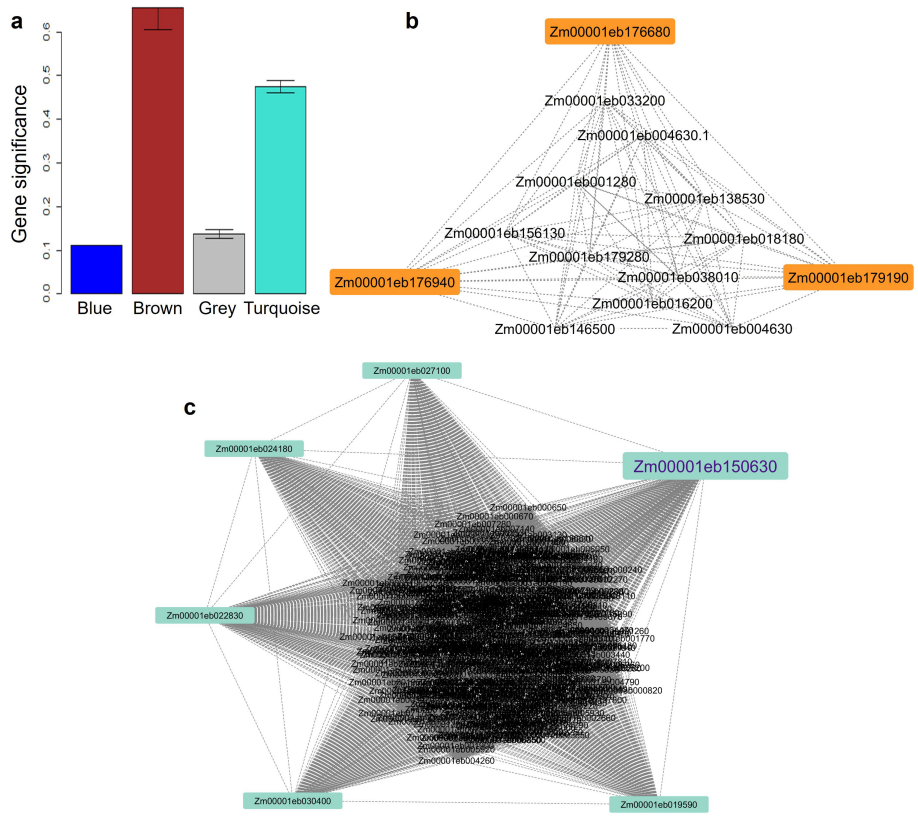


FIGURE 1
The gene significance values (*P*-value=8.9e-124) across co-expression network modules **(a)**. Brown **(b)** and Turquoise **(c)** significant modules showing three and six potential hub genes, respectively.

glucuronosyl/UDP-glucosyltransferase was found to be one of the 26 genes commonly regulated between maize, peanut, and cotton in response to *Aspergillus flavus*, and it was upregulated in both pericarp and seed tissues of cotton (Mehanathan et al., 2018).

Among the 12 most-significant genes predicted by SVM, RF, and GBM for abiotic stress, eight were upregulated under drought (Kakumanu et al., 2012; Song et al., 2017; Li et al., 2017; Thirunavukkarasu et al., 2017; SkZ et al., 2018; Yang et al., 2019; Danilevskaya et al., 2019; Zenda et al., 2019; Jia et al., 2020; Li et al., 2021; Kim et al., 2021; Maheswari et al., 2021; Bai et al., 2022), cold (Sobkowiak et al., 2014; Mao et al., 2017; Goering, 2017; Li et al., 2017; Avila et al., 2018; Waititu et al., 2021; Li et al., 2021), heat (Li et al., 2014; Shi et al., 2017; Li et al., 2017 and Zhao et al., 2019), and salt (Zhang et al., 2015; Du et al., 2017; Li et al., 2017; Chen et al., 2020; Zhang et al., 2021; 2022) stress of which four were downregulated under waterlogging condition (Supplementary Table 6) (Arora et al., 2017; Yao, 2021). Also, eight of the genes were commonly predicted between abiotic and combined stresses. Of the four genes specifically uniquely predicted for abiotic stress, *Zm00001eb012040* (GO:0006470) is expressed as a PPM-type phosphatase protein and was upregulated in all but waterlogging conditions whereas *Zm00001eb160470*, which also codes for a protein phosphatase, was upregulated under drought, heat and salt but downregulated under cold and waterlogging conditions (Supplementary Table 6). These genes are known to dephosphorylate serine/threonine in stress responsive genes in maize, thus impacting its response to multiple stresses (drought, salt) via hormone signal transductions (He et al., 2019). On the other hand, *Zm00001eb021010* (cysteine-rich and transmembrane domain-containing protein WIH1) was downregulated in all abiotic stresses except slight upregulation under waterlogging conditions. Several members of the cysteine-rich peptide family have been shown to respond extensively to various abiotic stresses in different plants including maize (Xu et al., 2018). *Zm00001eb050500* (GO:0009653), a top gene under abiotic as well as combined stress, encodes expansin, which plays an important role in stress relaxation of (arabino) xylan-cellulose networks within the cell wall (Yennawar et al., 2006). bZip-transcription factors (e.g., *Zm00001eb058820*) have been extensively studied for their critical roles in regulating plants response to abiotic stresses through morphological adaptations (Guo et al., 2024). In maize, a member of bZIP family positive regulated stress resistance through ABA-dependent signaling (He et al., 2024). *Zm00001eb044020* (a P-loop containing nucleoside triphosphate hydrolase), which was

moderately upregulated under cold and waterlogging conditions, but highly upregulated in response to drought, heat and salt, was also identified to be induced by proline under low water potential (Verslues et al., 2014) and was located within the genomic region associated with heat stress (Bashir et al., 2025). Only one gene *Zm00001eb038720*, expressed as Ribonuclease E/G, was found commonly predicted between biotic and abiotic stress, although it was not identified as one of the top-most significant genes by the two or more high-performing accurate models. Ribonuclease E/G are chloroplastic endoribonuclease that play crucial roles in plant's response to biotic and abiotic stresses because of their involvement in cleavage-mediated RNA homeostasis (Schein et al., 2008). The RNase E/G enzymes influence RNA modifications in plant adaptation to stresses by controlling mRNA stability and subsequent translation of genes (Cai et al., 2025).

Three genes that were commonly predicted as top-most candidates between abiotic stress and combined stresses as well as identified as the hub genes in the brown module, coded for a *bZIP* transcription factor 68 (*Zm00001eb176680*), glycine-rich cell wall structural protein 2 (*Zm00001eb176940*), and aldehyde dehydrogenase 11 (*ALDH11; Zm00001eb179190*). *bZIP* transcription factor 68 showed a negative response to cold stress in transgenic maize plants (Zeng et al., 2021). bZIP68 interacts with mitogen-activated protein kinase 8, which is also a negative regulator of the cold-stress response. A 358-bp indel in the *bZIP68* promoter region increased its expression resulting in decreased cold tolerance in maize (Li et al., 2022). *ZmbZIP4* was also shown to be differentially stimulated by high salinity, drought, heat, cold, and abscisic acid treatment in maize seedlings (Ma et al., 2018). Glycine-rich cell wall structural proteins (GRCWSPs) are known to be involved in plant's response to abiotic stress, especially under osmotic stress because of their roles in maintaining cell wall integrity during dehydration by connecting the lignin rings to strengthen the cell wall (Ryser et al., 2004). The GRCWSPs interact with cell-wall associated kinases to initiate recognition of environmental stimuli and subsequent signal transduction (Park et al., 2001). *ALDH11* showed upregulation under drought, cold, heat and salt and negative regulation under waterlogging conditions. Several *ALDH* genes have been reported to contribute to improving salt and drought tolerance in plants (Wang et al., 2024).

Promoter motifs in the top significant genes

Promoter regions corresponding to the top-most significantly predicted genes for different stresses showed 19 different CREs. As expected of the promoters, CAAT-box, TATA-box, and Unnamed_4 motifs existed in all promoters in high numbers (Supplementary Table 7). While there was no overall pattern in the distribution of the CREs among the promoters of the genes, the top biotic stress responsive genes, except *Zm00001eb042770*, had less overall total CREs. For most of the abiotic and combined stress responsive genes, the number of antioxidant responsive element (ARE) and ABRE (abscisic acid responsive element) were higher than the biotic stress responsive genes. The number of CREs were higher for two genes *Zm00001eb176940* (168) and *Zm00001eb179190* (169) predicted common between abiotic and combined stresses than the other genes. The ABRE and G-Box were found in more abundance in genes *Zm00001eb012040*, *Zm00001eb058820*, *Zm00001eb112930*, *Zm00001eb018700*, *Zm00001eb126900*, *Zm00001eb018700*, and *Zm00001eb126900* that were identified as top-most significant genes for abiotic and combined stresses.

Conclusion

The integration of a large volume of gene expression data from several RNA-seq studies with machine learning methods increased the generalizability and statistical power and allowed us to analyze the stress responses of the genes to identify a set of top-most genes with significant associations with (a)biotic stress in maize. The GO and KEGG pathway enrichment analysis of top genes provided clues to the mechanisms underlying maize's response to both biotic and abiotic stress conditions. Furthermore, the randomization procedure used in WGCNA method led to the identification of the hub genes validating their gene connectivity and possible interaction with other genes. Some of the genes identified in this study through ML were also found related to *Aspergillus flavus* resistance in maize in previous studies. Further functional validation of the roles of the 19 unique top-ranked genes including hub genes, predicted by the high-performing models, will lead to their utilization in developing multiple stress-resistant maize varieties.

Data availability statement

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

Author contributions

AP: Data curation, Formal Analysis, Software, Visualization, Writing – original draft. PG: Data curation, Formal Analysis, Methodology, Writing – review & editing. KR: Funding acquisition, Visualization, Writing – review & editing. NB: Conceptualization, Funding acquisition, Investigation, Methodology, Supervision, Visualization, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. The research was funded by the USDA-ARS (NACA Agreement # 58-6054-0-010) and USDA-NIFA Hatch grant (Accession # 7006617). The manuscript was accepted for publication by the Louisiana Agricultural Experiment Station as MS# 2025-306-40333.

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 handling editor GS declared a past co-authorship with the author AP.

Generative AI statement

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

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.2025.1611784/full#supplementary-material

References

Arora, K., Panda, K. K., Mittal, S., Mallikarjuna, M. G., Rao, A. R., Dash, P. K., et al. (2017). RNAseq revealed the important gene pathways controlling adaptive mechanisms under waterlogged stress in maize. *Sci. Rep.* 7, 10950. doi: 10.1038/s41598-017-10561-1

Avila, L. M., Obeidat, W., Earl, H., Niu, X., Hargreaves, W., and Lukens, L. (2018). Shared and genetically distinct Zea mays transcriptome responses to ongoing and past low temperature exposure. *BMC Genomics* 19, 761. doi: 10.1186/s12864-018-5134-7

Bai, M., Zeng, W., Chen, F., Ji, X., Zhuang, Z., Jin, B., et al. (2022). Transcriptome expression profiles reveal response mechanisms to drought and drought-stress mitigation mechanisms by exogenous glycine betaine in maize. *Biotechnol. Letters.* 44, 367–386. doi: 10.1007/s10529-022-03221-6

Baisakh, N., Da Silva, E. A., Pradhan, A. K., and Rajasekaran, K. (2023). Comprehensive meta-analysis of QTL and gene expression studies identify candidate genes associated with Aspergillus flavus resistance in maize. *Front. Plant Science.* 14. doi: 10.3389/fpls.2023.1214907

Bashir, L., Budhlakoti, N., Pradhan, A.K., Mehmood, A., Haque, M., Jacob, S. R, et al. (2025). Unraveling the genetic basis of heat tolerance and yield in bread wheat: QTN discovery and its KASP-assisted validation. *BMC Plant Biol.* 25 (268). doi: 10.1186/s12870-025-06285-4

Bedre, R., Rajasekaran, K., Mangu, V. R., Sanchez Timm, L. E., Bhatnagar, D., and Baisakh, N. (2015). Genome-wide transcriptome analysis of cotton (Gossypium hirsutum l.) identifies candidate gene signatures in response to aflatoxin producing fungus Aspergillus flavus. *PloS One* 10, e0138025. doi: 10.1371/journal.pone.0138025

Cai, J., Shen, L., Kang, H., and Xu, T. (2025). RNA modifications in plant adaptation to abiotic stresses. *Plant Communications*, 6(2). doi: 10.1016/j.xplc.2024.101229

Chen, F., Fang, P., Zeng, W., Ding, Y., Zhuang, Z., and Peng, Y. (2020). Comparing transcriptome expression profiles to reveal the mechanisms of salt tolerance and exogenous glycine betaine mitigation in maize seedlings. *PloS One* 15, e0233616. doi: 10.1371/journal.pone.0233616

Danilevskaya, O. N., Yu, G., Meng, X., Xu, J., Stephenson, E., Estrada, S., et al. (2019). Developmental and transcriptional responses of maize to drought stress under field conditions. *Plant Direct.* 3, e00129. doi: 10.1002/pld3.129

de Silva, K.K., Dunwell, J.M., and Wickramasuriya, A.M. (2022). Weighted gene correlation network Aaalysis (WGCNA) of arabidopsis somatic embryogenesis (SE) and identification of key gene modules to uncover SE-Associated hub genes. *Int. J. Genom.* 2022 (1), 7471063. doi: 10.1155/2022/7471063

Du, X., Wang, G., Ji, J., Shi, L., Guan, C., and Jin, C. (2017). Comparative transcriptome analysis of transcription factors in different maize varieties under salt stress conditions. *Plant Growth Regulation.* 81, 183–195. doi: 10.1007/s10725-016-0192-9

Eriksson, L., Johansson, E., Kettaneh-Wold, N., Trygg, J., Wikström, C., and Wold, S. (1999). Multi- and Megavariate Data Analysis: Principles and Applications. Umetrics Academy, Umea, Sweden, 533.

Farooqi, M. Q.U., Nawaz, G., Wani, S. H., Choudhary, J. R., Rana, M., Sah, R. P, et al. (2022). Recent developments in multi-omics and breeding strategies for abiotic stress tolerance in maize (Zea mays L.). *Frontiers in Plant Science*, 13, 965878. doi: 10.3389/fpls.2022.965878

Gharabli, H., Della Gala, V., and Welner, D. H. (2023). The function of UDP-glycosyltransferases in plants and their possible use in crop protection. *Biotechnol. Advances.* 67, 108182. doi: 10.1016/j.bioteChadv.2023.108182

Goering, R. N. (2017). Uncovering candidate cold tolerance genes in maize (*Zea mays*). *Departmental Honors Projects*. 54, 22–23. Available online at: https://digitalcommons.hamline.edu/dhp/54.

Guo, Z., Dzinyela, R., Yang, L., and Hwarari, D. (2024). bZIP transcription factors: Structure, modification, abiotic stress responses, and application in plant improvement. *Plants*, 13(15), 2058. doi: 10.3390/plants13152058

Han, G., Li, C., Xiang, F., Zhao, Q., Zhao, Y., Cai, R., et al. (2020). Genome-wide association study leads to novel genetic insights into resistance to *Aspergillus flavus* in maize kernels. *BMC Plant Biol.* 20, 1–11. doi: 10.1186/s12870-020-02404-5

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction (2nd ed.).* (New York: Springer). doi: 10.1007/978-0-387-84858-7

Hayford, R. K., Haley, O. C., Cannon, E. K., Portwood, II, , J. L., Gardiner, J. M., Andorf, C. M., and Woodhouse, M. R. (2024). Functional annotation and meta-analysis of maize transcriptomes reveal genes involved in biotic and abiotic stress. *BMC Genomics*, 25, 533. doi: 10.1186/s12864-024-10443-7

He, Z., Wu, J., Sun, X., and Dai, M. (2019). The maize clade A PP2C phosphatases play critical roles in multiple abiotic stress responses. *Int. J. Mol. Sci.* 20, 3573. doi: 10.3390/ijms20143573

He, S., Li, Y., Wang, Y., and Zhang, X. (2024). Molecular characterization of a stress-response bZIP transcription factor in banana. *Plant Cell, Tissue and Organ Culture*, 113 (2), 173–187. doi: 10.1007/s11240-013-0395-2

He, W., Zhu, Y., Leng, Y., Yang, L., Zhang, B., Yang, J., et al. (2021). Transcriptomic analysis reveals candidate genes responding maize gray leaf spot caused by *Cercospora zeina*. *Plants* 10, 2257. doi: 10.3390/plants10112257

Jia, S., Li, H., Jiang, Y., Tang, Y., Zhao, G., Zhang, Y., et al. (2020). Transcriptomic analysis of female panicles reveals gene expression responses to drought stress in maize (*Zea mays* L.). *Agronomy* 10, 313. doi: 10.3390/agronomy10020313

Kakumanu, A., Ambavaram, M. M., Klumas, C., Krishnan, A., Batlang, U., Myers, E., et al. (2012). Effects of drought on gene expression in maize reproductive and leaf meristem tissue revealed by RNA-Seq. *Plant Physiol.* 160, 846–867. doi: 10.1104/pp.112.200444

Kebede, A. Z., Johnston, A., Schneiderman, D., Bosnich, W., and Harris, L. J. (2018). Transcriptome profiling of two maize inbreds with distinct responses to *Gibberella* ear rot disease to identify candidate resistance genes. *BMC Genomics* 19, 1–12. doi: 10.1186/s12864-018-4513-4

Keel, B. N., and Lindholm-Perry, A. K. (2022). Recent developments and future directions in meta-analysis of differential gene expression in livestock RNA-Seq. *Front. Genet.* 13. doi: 10.3389/fgene.2022.983043

Kim, D., Paggi, J. M., Park, C., Bennett, C., and Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*, 37(8), 907–915. doi: 10.1038/s41587-019-0201-4

Kim, H., Lee, S., and Park, J. (2022). Identification of key genes in stress response using machine learning models. Journal of Computational Biology, 29(4), 345–358. doi: 10.1089/cmb.2022.0123

Kim, K. H., Song, K., Park, J. M., Kim, J. Y., and Lee, B. M. (2021). RNA-Seq analysis of gene expression changes related to delay of flowering time under drought stress in tropical maize. *Appl. Sci.* 11, 4273. doi: 10.3390/app11094273

Kimotho, R. N., Baillo, E. H., and Zhang, Z. (2019). Transcription factors involved in abiotic stress responses in Maize (*Zea mays* L.) and their roles in enhanced productivity in the post genomics era. *PeerJ* 7, e7211. doi: 10.7717/peerj.7211

Kira, K., and Rendell, A. L. (1992). "The feature selection problem: Traditional methods and a new algorithm," in *Proceedings of the tenth national conference on Artificial intelligence (AAAI'92)*. (San Jose, CA: AAAI Press) 129–134.

Lambarey, H., Moola, N., Veenstra, A., Murray, S., and Suhail Rafudeen, M. (2020). Transcriptomic analysis of a susceptible African maize line to Fusarium verticillioides infection. *Plants* 9, 1112. doi: 10.3390/plants9091112

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinf.* 9, 1–13. doi: 10.1186/1471-2105-9-559

Li, P., Cao, W., Fang, H., Xu, S., Yin, S., Zhang, Y., et al. (2017). Transcriptomic profiling of the maize (*Zea mays* L.) leaf response to abiotic stresses at the seedling stage. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.00290

Li, Y., Hu, J., Liu, J., Suo, H., Yu, Y., and Han, F. (2014). Genome-wide analysis of gene expression profiles during early ear development of sweet corn under heat stress. *Plant Breeding*, 134(1), 17–27. doi: 10.1111/pbr.12235

Li, Z., Fu, D., Wang, X., Zeng, R., Zhang, X., Tian, J., et al. (2022). The transcription factor bZIP68 negatively regulates cold tolerance in maize. *Plant Cell.* 34, 2833–2851. doi: 10.1093/plcell/koac137

Li, H., Yue, H., Xie, J., Bu, J., Li, L., Xin, X., et al. (2021). Transcriptomic profiling of the high-vigour maize (*Zea mays* L.) hybrid variety response to cold and drought stresses during seed germination. *Sci. Rep.* 11, 19345. doi: 10.1038/s41598-021-98907-8

Liao, Y., Smyth, G. K., and Shi, W. (2014). featureCounts: An efficient general-purpose program for assigning sequence reads to genomic features. *Bioinformatics*, 30 (7), 923–930. doi: 10.1093/bioinformatics/btt656

Liu, H., Wu, H., Wang, Y., Wang, H., Chen, S., and Yin, Z. (2021). Comparative transcriptome profiling and co-expression network analysis uncover the key genes associated with early-stage resistance to *Aspergillus flavus* in maize. *BMC Plant Biol.* 21, p.216. doi: 10.1186/s12870-021-02983-x

Ma, H., Liu, C., Li, Z., Ran, Q., Xie, G., Wang, B., et al. (2018). ZmbZIP4 contributes to stress resistance in maize by regulating ABA synthesis and root development. *Plant Physiol.* 178, 753–770. doi: 10.1104/pp.18.00436

Mahendran, N., Durai Raj Vincent, P. M., Srinivasan, K., and Chang, C. Y. (2020). Machine learning based computational gene selection models: a survey, performance evaluation, open issues, and future research directions. *Front. Genet.* 11. doi: 10.3389/fgene.2020.603808

Maheswari, M., Varalaxmi, Y., Sarkar, B., Ravikumar, N., Vanaja, M., Yadav, S. K., et al. (2021). Tolerance mechanisms in maize identified through phenotyping and transcriptome analysis in response to water deficit stress. *Physiol. Mol. Biol. Plants.* 27, 1377–1394. doi: 10.1007/s12298-021-01003-4

Mallikarjuna, M. G., Thirunavukkarasu, N., Sharma, R., Shiriga, K., Hossain, F., Bhat, J. S., et al. (2020). Comparative transcriptome analysis of iron and zinc deficiency in maize (*Zea mays* L.). *Plants* 9, 1812. doi: 10.3390/plants9121812

Mao, J., Yu, Y., Yang, J., Li, G., Li, C., Qi, X., et al. (2017). Comparative transcriptome analysis of sweet corn seedlings under low-temperature stress. *Crop J.* 5, 396–406. doi: 10.1016/j.cj.2017.03.005

Meale, A. D., Blomstedt, C. K. M., Hamill, J. D., Gaff, D. F., Griffiths, C., Islam, S., et al. (2015). "Method for improving crop productivity," in *International Application Published Under The Patent Cooperation Treaty (PCT)*.

Mehanathan, M., Bedre, R., Mangu, V., Rajasekaran, K., Bhatnagar, D., and and Baisakh, N. (2018). Identification of candidate resistance genes of cotton against Aspergillus flavus infection using a comparative transcriptomics approach. *Physiology and Molecular Biology of Plants*, 24(3), 513–519. doi: 10.1007/s12298-018-0522-7

Musungu, B., Bhatnagar, D., Quiniou, S., Brown, R. L., Payne, G. A., O'Brian, G., et al. (2020). Use of dual RNA-seq for systems biology analysis of *Zea mays* and *Aspergillus flavus* interaction. *Front. Microbiol.* 11. doi: 10.3389/fmicb.2020.00853

Nazari, L., Aslan, M. F., Sabanci, K., and Ropelewska, E. (2023). Integrated transcriptomic meta-analysis and comparative artificial intelligence models in maize under biotic stress. *Sci. Rep.* 13, 15899. doi: 10.1038/s41598-023-42984-4

Nazari, L., Zinati, Z., and Bagnaresi, P. (2024). Identification of biomarker genes from multiple studies for abiotic stress in maize through machine learning. *Journal of Biosciences*, 49(1), 1. doi: 10.1007/s12038-023-00392-w

Park, A., Cho, R., , S., Yun, K., , U., Jin, J., , M., Lee, Y., , S., Sachetto-Martins, H., Park, G., , O., , K., et al. (2001). Interaction of the Arabidopsis receptor protein kinase Wak1 with a glycine-rich protein, AtGRP-3. *Journal of Biological Chemistry*, 276(28), 26688–26693. doi: 10.1074/jbc.M101283200

Pérez-Enciso, M., and Tenenhaus, M. (2003). Prediction of clinical outcome with microarray data: a partial least squares discriminant analysis (PLS-DA) approach. *Hum. Genet.* 112, 581–592. doi: 10.1007/s00439-003-0921-9

Qin, F., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2011). Achievements and challenges in understanding plant abiotic stress responses and tolerance. *Plant Cell Physiol.* 52, 1569–1582. doi: 10.1093/pcp/pcr106

Ramasamy, A., Mondry, A., Holmes, C. C., and Altman, D. G. (2008). Key issues in conducting a meta-analysis of gene expression microarray datasets. *PloS Med.* 5, e184. doi: 10.1371/journal.pmed.0050184

Ramegowda, V., and Senthil-Kumar, M. (2015). The interactive effects of simultaneous biotic and abiotic stresses on plants: mechanistic understanding from drought and pathogen combination. *J. Plant Physiol.* 176, 47–54. doi: 10.1016/j.jplph.2014.11.008

Ryser, U., Schorderet, M., Guyot, R., and Keller, B. (2004). A new structural element containing glycine-rich proteins and rhamnogalacturonan I in the protoxylem of seed plants. *Journal of Cell Science*, 117(7), 1179–1190. doi: 10.1242/jcs.00966

Sabanci, K., Aslan, M. F., Ropelewska, E., Unlersen, M. F., and Durdu, A. (2022). A novel convolutional-recurrent hybrid network for sunn pest–damaged wheat grain detection. *Food Analytical Methods*, 15, 1748–1760. doi: 10.1007/s12161-022-02251-0

Schadt, E. E., Li, C., Ellis, B., and Wong, W.H. (2001). Feature extraction and normalization algorithms for high-density oligonucleotide gene expression array data. *Journal of Cellular Biochemistry*, 80(2), 192–202. doi: 10.1002/jcb.10073

Schurack, S., Depotter, J. R., Gupta, D., Thines, M., and Doehlemann, G. (2021). Comparative transcriptome profiling identifies maize line specificity of fungal effectors in the maize–*Ustilago maydis* interaction. *Plant J.* 106, 733–752. doi: 10.1111/tpj.15195

Schein, A., Sheffy-Levin, S., Glaser, F., and Schuster, G. (2008). The RNase E/G-type endoribonuclease of higher plants is located in the chloroplast and cleaves RNA similarly to the E. coli enzyme. *RNA*, 14(6), 1057–1068. doi: 10.1261/rna.907608

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303

Sharma, R., De Vleesschauwer, D., Sharma, M. K., and Ronald, P. C. (2013). Recent advances in dissecting stress-regulatory crosstalk in rice. *Mol. Plant* 6, 250–260. doi: 10.1093/mp/sss147

Shi, J., Yan, B., Lou, X., Ma, H., and Ruan, S. (2017). Comparative transcriptome analysis reveals the transcriptional alterations in heat-resistant and heat-sensitive sweet maize (Zea mays L.) varieties under heat stress. *BMC Plant Biol.* 17, 1–10. doi: 10.1186/s12870-017-0973-y

Shi, F., Zhang, Y., Wang, K., Meng, Q., Liu, X., Ma, L., et al. (2018). Expression profile analysis of maize in response to *Setosphaeria turcica*. *Gene* 659, 100–108. doi: 10.1016/j.gene.2018.03.030

Shu, X., Livingston, D. P. III, Woloshuk, C. P., and Payne, G. A. (2017). Comparative histological and transcriptional analysis of maize kernels infected with *Aspergillus flavus* and *Fusarium verticillioides*. *Front. Plant Science.* 8. doi: 10.3389/fpls.2017.02075

SkZ, A., Vardharajula, S., and Vurukonda, S. S. K. P. (2018). Transcriptomic profiling of maize (*Zea mays* L.) seedlings in response to *Pseudomonas putida* stain FBKV2 inoculation under drought stress. *Ann. Microbiol.* 68, 331–349. doi: 10.1007/s13213-018-1341-3

Sobkowiak, A., Jończyk, M., Jarochowska, E., Biecek, P., Trzcinska-Danielewicz, J., Leipner, J., et al. (2014). Genome-wide transcriptomic analysis of response to low temperature reveals candidate genes determining divergent cold-sensitivity of maize inbred lines. *Plant Mol. Biol.* 85, 317–331. doi: 10.1007/s11103-014-0187-8

Song, K., Kim, H. C., Shin, S., Kim, K. H., Moon, J. C., Kim, J. Y., et al. (2017). Transcriptome analysis of flowering time genes under drought stress in maize leaves. *Front. Plant Science.* 8. doi: 10.3389/fpls.2017.00267

Thirunavukkarasu, N., Sharma, R., Singh, N., Shiriga, K., Mohan, S., Mittal, S., et al. (2017). Genome wide expression and functional interactions of genes under drought stress in maize. *Int. J. Genomics* 2017, .2568706. doi: 10.1155/2017/2568706

Van Aken, O. (2008). Methods and means for the production of plants with improved stress resistance.

Verslues, P. E., Lasky, J. R., Juenger, T. E., Liu, T. W., and Kumar, M. N. (2014). Genome-wide association mapping combined with reverse genetics identifies new effectors of low water potential-induced proline accumulation in Arabidopsis. *Plant Physiology*, 164(1), 144–159. doi: 10.1104/pp.113.224014

Waititu, J. K., Cai, Q., Sun, Y., Sun, Y., Li, C., Zhang, C., et al. (2021). Transcriptome profiling of maize (*Zea mays* L.) leaves reveals key cold-responsive genes, transcription factors, and metabolic pathways regulating cold stress tolerance at the seedling stage. *Genes.* 12, 1638. doi: 10.3390/genes12101638

Wang, Y., Guo, H., Wu, X., Wang, J., Li, H., and Zhang, R. (2022). Transcriptomic and physiological responses of contrasting maize genotypes to drought stress. *Front. Plant Science.* 13. doi: 10.3389/fpls.2022.928897

Wang, J., Xing, C., Wang, H., Zhang, H., Wei, W., Xu, J., et al. (2024). Identification of key modules and hub genes involved in regulating the feather follicle development of Wannan chickens using WGCNA. *Poultry Science.* 103, 103903. doi: 10.1016/j.psj.2024.103903

Wetterhorn, K. M. (2018). *Enzymatic inactivation of Trichothecene mycotoxins associated with* Fusarium *head blight* (Madison, WI: The University of Wisconsin-Madison).

Wold, S., Martens, H., and Wold, H. (1983). The multivariate calibration problem in chemistry solved by the PLS method. In A. Ruhe & B. Kågström (Eds.), Proceedings of

the Conference on Matrix Pencils, 286–293. Springer-Verlag, Heidelberg. doi: 10.1007/978-3-642-61794-2_29

Xu, Y., Yu, Z. P., Zhang, D., Huang, J. G., Wu, C. G., Yang, G. D, et al. (2018). CYSTM, a novel non-secreted cysteine-rich peptide family, involved in environmental stresses in Arabidopsis thaliana. *Plant Cell Physiology*, 59(2), 423–438. doi: 10.1093/pcp/pcx204

Yang, M., Geng, M., Shen, P., Chen, X., Li, Y., and Wen, X. (2019). Effect of post-silking drought stress on the expression profiles of genes involved in carbon and nitrogen metabolism during leaf senescence in maize (Zea mays L.). *Plant Physiology and Biochemistry*, 135, 304–309. doi: 10.1016/j.plaphy.2018.12.025

Yao, Q. (2021). Crucial waterlogging-responsive genes and pathways revealed by comparative physiology and transcriptome in tropical and temperate maize (*Zea mays* L.) inbred lines. *J. Plant Biol.* 64, 313–325. doi: 10.1007/s12374-021-09298-2

Yennawar, N. H., Li, L.-C., Dudzinski, D. M., Tabuchi, A., and Cosgrove, D. J. (2006). Crystal structure and activities of EXPB1 (Zea m 1), a β-expansin and group-1 pollen allergen from maize. Proceedings of the National Academy of Sciences, 103 (40), 14664–14671. doi: 10.1073/pnas.0605979103

Zenda, T., Liu, S., Wang, X., Liu, G., Jin, H., Dong, A., et al. (2019). Key maize drought-responsive genes and pathways revealed by comparative transcriptome and physiological analyses of contrasting inbred lines. *Int. J. Mol. Sci.* 20, 1268. doi: 10.3390/ijms20061268

Zeng, R., Li, Z., Shi, Y., Fu, D., Yin, P., Cheng, J., et al. (2021). Natural variation in a type-A response regulator confers maize chilling tolerance. *Nat. Commun.* 12, 4713. doi: 10.1038/s41467-021-25001-y

Zhang, M., Kong, X., Xu, X., Li, C., Tian, H., and Ding, Z. (2015). Comparative transcriptome profiling of the maize primary, crown and seminal root in response to salinity stress. *PloS One* 10, e0121222. doi: 10.1371/journal.pone.0121222

Zhang, X., Liu, J., Huang, Y., Wu, H., Hu, X., Cheng, B., et al. (2022). Comparative transcriptomics reveal the molecular mechanism of the parental lines of maize hybrid An'nong876 in response to salt stress. *Int. J. Mol. Sci.* 23, 5231. doi: 10.3390/ijms23095231

Zhang, X., Liu, P., Qing, C., Yang, C., Shen, Y., and Ma, L. (2021). Comparative transcriptome analyses of maize seedling root responses to salt stress. *PeerJ* 9, e10765. doi: 10.7717/peerj.10765

Zhao, Y., Hu, F., Zhang, X., Wei, Q., Dong, J., Bo, C., et al. (2019). Comparative transcriptome analysis reveals important roles of nonadditive genes in maize hybrid An'nong 591 under heat stress. *BMC Plant Biol.* 19, 1–17. doi: 10.1186/s12870-019-1878-8

Zhu, J. K. (2002). Salt and drought stress signal transduction in plants. *Annu. Rev. Plant Biol.* 53, 247. doi: 10.1146/annurev.arplant.53.091401.143329

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 21 July 2025
DOI 10.3389/fpls.2025.1633325

OPEN ACCESS

EDITED BY
Gurjeet Singh,
Texas A and M University, United States

REVIEWED BY
Dandan Hu,
Henan Agricultural University, China
Francia Ravelombola,
University of Missouri, United States

*CORRESPONDENCE
Bixian Zhang
hljsnkyzbx@163.com
Honglei Ren
renhonglei2022@163.com

†These authors have contributed equally to this work

RECEIVED 22 May 2025
ACCEPTED 04 July 2025
PUBLISHED 21 July 2025

CITATION
Zhang C, Wang Y, Zhang R, Yuan R, Zhao K, Liu X, Wang X, Zhang F, Lamlom SF, Zhang B and Ren H (2025) Comprehensively characterize the soybean CAM/CML gene family, as it provides resistance against both the soybean mosaic virus and *Cercospora sojina* pathogens. *Front. Plant Sci.* 16:1633325. doi: 10.3389/fpls.2025.1633325

Comprehensively characterize the soybean CAM/CML gene family, as it provides resistance against both the soybean mosaic virus and *Cercospora sojina* pathogens

Chunlei Zhang[1†], Yanbo Wang[1,2†], Ruiping Zhang[1], Rongqiang Yuan[1], Kezhen Zhao[1], Xiulin Liu[1], Xueyang Wang[1], Fengyi Zhang[1], Sobhi F. Lamlom[1,3], Bixian Zhang[4*] and Honglei Ren[1*]

[1]Soybean Research Institute of Heilongjiang Academy of Agriculture Sciences, Harbin, China, [2]College of Life Science, Northeast Agriculture University, Harbin, China, [3]Plant Production Department, Faculty of Agriculture Saba Basha, Alexandria University, Alexandria, Egypt, [4]Institute of Biotechnology of Heilongjiang Academy of Agricultural Sciences, Harbin, China

Introduction: Calmodulin (CAM) and calmodulin-like (CML) proteins are essential calcium sensors that mediate plant responses to biotic and abiotic stresses. In soybean (Glycine max L. Merr.), these proteins play critical roles in resistance to multiple pathogens, yet a comprehensive characterization of this gene family and its connection to disease resistance has been lacking.

Methods: This study identified and characterized 113 CAM/CML genes in the soybean genome, including 11 GmCAMs and 102 GmCMLs, through bioinformatic analysis using sequence homology, domain architecture, and phylogenetic approaches. Gene structure analysis, cis-acting element identification, and expression profiling were conducted to examine functional diversification and pathogen response patterns.

Results: Phylogenetic analysis revealed 14 distinct groups with evidence of both ancient and recent gene duplication events contributing to family expansion. Gene structure analysis demonstrated higher conservation among GmCAMs (with all but one containing introns) compared to GmCMLs (70% intronless). Analysis of cis-acting elements indicated enrichment of hormone-responsive elements, particularly those associated with abscisic acid (31.2%) and methyl jasmonate (27.7%) responses. Expression profiling revealed distinct CAM/CML gene expression patterns in response to two major soybean pathogens: Soybean Mosaic Virus (SMV) and Cercospora sojina. We identified 15 GmCAM/CML genes that exhibited significantly altered expression in response to both pathogens, with GmCML23, GmCML47, and GmCAM4 showing the strongest correlation with resistance phenotypes.

Discussion: The expression patterns of these genes were validated in various resistant and susceptible varieties, confirming their potential role in broad-

spectrum disease resistance. Our findings offer valuable insights into the evolutionary history and functional diversification of the soybean CAM/CML gene family and identify promising candidates for enhancing soybean resistance to multiple pathogens through molecular breeding strategies.

KEYWORDS

soybean mosaic virus (SMV), *C. sojina*, soybean, CAM/CML genes, disease resistance

1 Introduction

Plants encounter a variety of biotic and abiotic stresses throughout their growth and development, which can considerably diminish yields. They address these challenges through complex internal signaling pathways. Soybean plants (*Glycine max* L. Merr.) constantly encounter multiple challenges that can significantly affect growth, development, and yield (Majidian et al., 2024; Ren et al., 2024). As immobile organisms, soybeans are continually exposed to various environmental stimuli and pathogenic microorganisms, including bacteria, fungi, and viruses, all of which can lead to substantial crop losses (Ren et al., 2023). To survive these adverse conditions, soybeans have evolved sophisticated molecular mechanisms for perceiving external signals and translating them into appropriate cellular responses (Zhao et al., 2013).

Among the most economically significant pathogens affecting soybean production are Soybean Mosaic SMV and *C. sojina*, the causal agent of gray leaf spot (GLS) (Madhusudhan; Barro et al., 2023). SMV, a member of the Potyvirus genus, represents one of the most widespread and damaging viral diseases in soybean cultivation worldwide, causing yield losses ranging from 8% to 94% depending on cultivar susceptibility, infection timing, and environmental conditions (Choi et al., 2005; Zhang et al., 2009a). The virus is transmitted through aphid vectors and infected seeds, making it particularly challenging to control through conventional management practices. Similarly, *C. sojina* causes gray leaf spot, a foliar disease that has emerged as a major threat to soybean production, particularly in warm, humid environments (Kim et al., 2013). This fungal pathogen can cause premature defoliation, reduced photosynthetic capacity, and significant yield reductions of up to 60% in susceptible cultivars under favorable disease conditions (Yang and Luo, 2021). The increasing prevalence of both pathogens, combined with their potential for rapid spread and adaptation, underscores the critical need for understanding the molecular mechanisms underlying soybean resistance to these diseases (Jeena et al., 2024).

Calcium ions (Ca^{2+}) serve as universal secondary messengers in these signaling networks, playing crucial roles in both regulating plant growth and mediating adaptations to biotic and abiotic stresses (Tong et al., 2021; Li et al., 2022). When environmental threats are detected, a rapid increase in cytoplasmic calcium concentration occurs, resulting in calcium transients and oscillations that represent the plant's initial response to stimuli. Specialized calcium-binding proteins then decode these calcium signatures (Tuteja and Mahajan, 2007). In Arabidopsis, over 250 calcium sensor proteins have been identified, including calcineurin B-like proteins (*CBLs*), CaM, CMLs, calcium-dependent protein kinases (*CPKs*), and calcium and calmodulin-dependent protein kinase (*CCaMK*), and all of them contain different numbers of EF-hand motifs (Cheng et al., 2002; Day et al., 2002; Yang and Poovaiah, 2003; Luan, 2009; Boudsocq and Sheen, 2013; Wang et al., 2015).

CaMs/CMLs are essential types of Ca^{2+} sensors and are crucial components in Ca^{2+} signal transduction (Perochon et al., 2011). CaMs, which contain four EF-hand motifs, are conserved Ca^{2+} sensors found in both plants and animals (Mccormack and Braam, 2003). On the other hand, CMLs, which typically contain 1–6 EF-hand motifs, show some sequence similarity to CaM and display structural variations in plants (Gifford et al., 2007). Genome-wide identification and analysis of CaM/CML genes have been conducted for numerous plant species, including *Arabidopsis* (7 *CaMs* and 50 *CMLs*), rice (*Oryza sativa*, 5 *CaMs* and 32 *CMLs*), and *Brassica napus* (25 *CaMs* and 168 *CMLs*) (Yang and Poovaiah, 2003; Boonburapong and Buaboocha, 2007; He et al., 2020).

While CMLs and CaMs are homologous, plants have significantly larger CMLs than CaMs. The roles of CaMs and CMLs in stress response are well established. For instance, the AtCaM3 knockout mutant in Arabidopsis exhibits decreased heat tolerance, while transgenic lines overexpressing AtCaM3 demonstrate improved heat tolerance (Zhang et al., 2009b). *AtCML8* and *AtCML9* enhance Arabidopsis resistance to *Pseudomonas syringae* via the ABA and SA pathways (Zhu et al., 2017). The *AtCML24* gene plays a role in inhibiting pathogen-induced nitric oxide (NO) generation (Ma et al., 2008). In cotton, *GhCML11* interacts with *GhMYB108*, acting as a positive regulator in defense against *Verticillium dahliae* infection (Cheng et al., 2016). In the soybean genome, at least 262 genes encode proteins containing one to six EF-hand motifs, including 6 CaMs, 144 CMLs, 15 calcineurin B-like proteins, 50 calcium-dependent protein kinases (*CDPKs*), and various other calcium-responsive elements (Zeng et al., 2017; Ramlal et al., 2024). These proteins contain conserved EF-hand domains—helix-loop-helix structures that bind calcium and undergo conformational changes to activate downstream signaling pathways. CaM is evolutionarily conserved

across eukaryotes and is the primary calcium sensor (Mohanta et al., 2017). Most soybean EF-hand genes (87.8%) contain at least one type of hormone signaling or stress response-related regulatory element in their promoter regions, indicating their potential involvement in stress adaptation mechanisms (Zeng et al., 2017). This extensive calcium-signaling network reflects the critical importance of these pathways in soybean's ability to respond to the environment (Kaur et al., 2022).

Calcium signaling plays a significant role in soybean defense against pathogenic microorganisms (Negi et al., 2023). Specific calmodulin isoforms from soybean, such as *SCaM-4* and *SCaM-5*, are rapidly induced during plant defense responses and can enhance resistance to a wide spectrum of pathogens, including bacteria, fungi, and viruses (Ramlal et al., 2024). These specialized CaM variants appear to activate salicylic acid-independent pathways leading to disease resistance (Gao et al., 2014; Arfaoui et al., 2016). The soybean *GmCaM4* gene has been extensively studied in this context. Overexpression of *GmCaM-4* and *GmCaM-5*, two divergent calmodulin isoforms from soybean, can induce expression of pathogenesis-related (PR) genes and enhance disease resistance (Park et al., 2004). This activation depends on NIM1 (Non-immunity 1), a key regulator of systemic acquired resistance, demonstrating the integration of calcium signaling with established plant immune pathways (Park et al., 2020; Vidhyasekaran and Vidhyasekaran, 2020). Beyond pathogen defense, calcium signaling networks help soybeans adapt to abiotic stresses such as drought, salinity, and temperature extremes (Park et al., 2004). Osmotic stress induces a series of molecular and cellular responses beginning with increased cytosolic calcium concentration, which subsequently activates appropriate cellular mechanisms to mitigate potential damage (Ramlal et al., 2024).

For instance, the transcription factor *AtMYB2*, which regulates salt and dehydration-responsive genes, was identified as a calmodulin-binding protein (Zeng et al., 2015). The salt-inducible soybean CaM isoform *ScaM4* increases *AtMYB2's* DNA binding activity, enhancing transcription of genes involved in proline synthesis and conferring salt tolerance (Ranty et al., 2006). This demonstrates how calcium signaling directly influences metabolic adaptations to environmental stress.

This study aims to thoroughly analyze the resistance mechanisms of the soybean disease resistance gene family to mosaic virus and *C. sojina*. By referencing published data on mosaic virus resistance and combining it with our own research on gray leaf spot, several significant gene loci related to disease resistance were identified, providing both theoretical foundations and practical guidance for breeding soybean disease resistance.

2 Materials and methods

2.1 Identification of CAM/CML genes in soybean

The soybean genome and GFF3 annotation files were obtained from the Soybase website. Seven *ATCAM* and fifty Arabidopsis *ATCML* proteins from the TAIR database were searched against protein sequences from the soybean genome database using BLASTP (E value < 1E-6). Additionally, hidden Markov model (HMM) profiles of the EF-hand domain (PF13499, PF13405, PF13202, PF13833) were downloaded from the PFAM database and utilized for HMM analysis with the TBtools tool. The results of BLASTP and HMM were compared and manually analyzed. The CAM/CML proteins were further evaluated using SMART (http://smart.embl-heidelberg.de/) and Interpro (http://www.ebi.ac.uk/interpro/), confirming the presence and integrity of the EF-hand domain without any other domains. The molecular weight (Mw) and isoelectric point (pI) of the *GmCAM/CML* proteins were predicted using the Protein Parameter Calc tool in TBtools.

2.2 Chromosome location, gene structure and conserved motif analysis

The chromosomal positions of CAM/CML genes were established by downloading the GFF file for the soybean genome and the CAM/CML family members from the Phytozome website using TBtools (Chen et al., 2023). Collinearity analysis of the CAM/CML genes was carried out with the one-step MCScanX plug-in of TBtools, and the results were visualized using the Advanced Circos plug-in (Chen et al., 2023). The Ka/Ks calculator in TBtools was utilized to compute nonsynonymous and synonymous substitution rates (Chen et al., 2023). The conserved *GmCAM/CML* protein motifs were predicted using the MEME program with default parameters, with a maximum of 5 motifs.

2.3 Multiple sequence alignment, phylogenetic analysis, and collinearity analysis

The Arabidopsis CAM/CML protein sequences (7 ATCAMS, 50 ATCMLS) were downloaded from the TAIR website. Multiple sequence alignments were conducted using the ClustalW tool. A rooted neighbor-joining phylogenetic (NJ) tree was constructed with Mega 11 software based on the full-length CAM/CML protein sequences to study the evolutionary relationships among CAM/CML proteins. The collinearity of CAM/CML genes was analyzed using the multiple scanning tool package MCSCANX, and the collinearity between homologous proteins in the Arabidopsis and rice genomes was examined. The results were visualized with TBTools. The KA/KS ratios among GmCAM/CML members were calculated using Ka/Ks_calculator2.0 software.

2.4 Prediction of cis-acting elements in the promoters of CAM/CML genes in soybean

We downloaded the promoter sequences (approximately 2 kb upstream of the transcription start sites) of soybean GmCAM/CML family genes from Phytozome, then used the PlantCARE online

website (http://bioinformatics.psbugent.be/webtools/plantcare/html/) to predict and analyze promoter elements in these genes, remove the core elements (TATA-box and CAAT-box) of the promoter, and visualize them using TBtools (Chen et al., 2023).

2.5 Abiotic and biotic stress treatments

Five resistant soybean varieties (24JD21, 24JD210, 24JD829, 24JD890, 24JD697) and five susceptible (24JD892, 24JD878, 24JD876, 24JD882, 24JD867) were selected for biotic stress treatments based on their response to *C. sojina* infection. The *C. sojina* isolate (strain *CS-2018*) was obtained from the National Soybean Improvement Center in Nanjing, China, and cultured on potato dextrose agar (PDA) medium at 25°C for 14 days. Spores were harvested by washing the plates with sterile water containing 0.05% Tween-20, and the concentration was adjusted to 5×10^5 spores/ml. Soybean plants were grown in a greenhouse under controlled conditions (25°C day/22°C night, 16-hour photoperiod, 70% relative humidity). Plants at the V3 stage (with the third trifoliate leaf fully developed) were used for inoculation. For *C. sojina* inoculation, the spore suspension was sprayed onto the plants until the runoff. The inoculated plants were kept in a humidity chamber for 48 hours and then returned to normal greenhouse conditions. Disease severity was assessed 14 days after inoculation using a 0–9 scale, where 0 = no lesions and 9 = severe infection with coalescing lesions covering more than 50% of the leaf area. Varieties with average scores of 0–3 were classified as resistant, 4–6 as moderately susceptible, and 7–9 as highly susceptible (Supplementary Table S1). For SMV response data, we utilized published transcriptome data from the comparative transcriptome analysis of early resistance of *Glycine max* to soybean mosaic virus (Li et al., 2023).

2.6 RNA extraction and RT-qPCR analyses

Total RNA was extracted using the RNA Extraction Kit from CWBIO (Jiangsu, China). This total RNA was reverse-transcribed to generate first-strand cDNA with the HiScript III RT SuperMix for qPCR (+gDNA wiper) provided by Vazyme (Beijing, China). Quantitative real-time PCR was conducted using a Roche instrument with ChamQ SYBR qPCR Master Mix from Vazyme (Beijing, China). The internal reference gene utilized was soybean β-Tubulin, and the relative expression levels of the target gene were calculated with the 2^-ΔΔCT method. In the expression profile analysis, significance was evaluated via the Student t-test, with levels of significance indicated by asterisks (*P < 0.05, **P < 0.01, and *** P < 0.001). The sequences of primers used in this study are listed in Supplementary Table S2. For the *C. sojina* infection time course, leaf samples were gathered 14 days after inoculation. Each variety had three biological replicates, with each replicate comprising the third trifoliate leaves from three plants. The collected leaf tissues were promptly frozen in liquid nitrogen and stored at -80°C until RNA extraction.

3 Results

3.1 Identification and annotation of CAM/CML gene family members in soybean

To identify CAM/CML genes in soybean, we employed a dual screening approach. First, we conducted BLASTP searches against the soybean genome using protein sequences of seven *ATCAMs* and fifty *ATCMLs* from Arabidopsis as queries, with an e-value threshold of 1 × 10^-6. Concurrently, we performed hidden Markov model (HMM) analyses by aligning AMP domain files (PF13499, PF13405, PF13202, PF13833) with all soybean amino acid sequences. Candidate genes identified through these methods were further verified using multiple domain databases including PFAM, SMART, and NCBI's Conserved Domain Database (CDD). This comprehensive screening approach resulted in the identification of 11 *GmCAM* and 102 *GmCML* genes. These genes were systematically named according to their chromosomal positions (*GmCAM1-GmCAM11* and *GmCML1-GmCML102*). Subsequent characterization of the physical and chemical properties of these CAM/CML family members revealed considerable diversity. The encoded proteins ranged in length from 74 to 371 amino acids, with theoretical molecular weights between 8622.56 and 43574.75 kDa. The isoelectric points (pI) of these proteins exhibited a broad distribution from 3.94 to 9.99, indicating diverse charge properties within this gene family. Detailed information for individual family members is provided in Supplementary Table S3.

3.2 Phylogenetic relationships reveal distinct evolutionary groups within the CAM/CML gene family

To investigate the evolutionary relationships among CAM/CML proteins, we constructed a neighbor-joining phylogenetic tree using maximum likelihood methods with 1,000 bootstrap replicates. The analysis included 57 Arabidopsis proteins (7 *AtCAMs* and 50 *AtCMLs*) and 113 soybean proteins (11 GmCAMs and 102 GmCMLs) (Figure 1). The phylogenetic analysis resolved the CAM/CML superfamily into 14 distinct monophyletic groups with strong bootstrap support (>70%). Group I emerged as the largest clade, containing 36 members (8 *AtCMLs* and 28 *GmCMLs*), suggesting an ancient expansion of CML proteins predating the Arabidopsis-soybean divergence approximately 54 million years ago. Groups II and III collectively harbored all CAM proteins, with Group II containing 15 members (3 *AtCMLs*, 9 *GmCMLs*, and 3 *GmCAMs*) and Group III containing 21 members (4 *AtCMLs*, 7 *AtCAMs*, 8 *GmCAMs*, and 2 *GmCMLs*). The exclusive clustering of CAM proteins within these two groups indicates their evolutionary origin from a common ancestral lineage and subsequent conservation of core calcium-sensing functions. The remaining groups (IV-XIV) comprised exclusively CML proteins, with variable sizes ranging from 2 to 22 members. Notably, Groups IV and VII represented the smallest clades, each

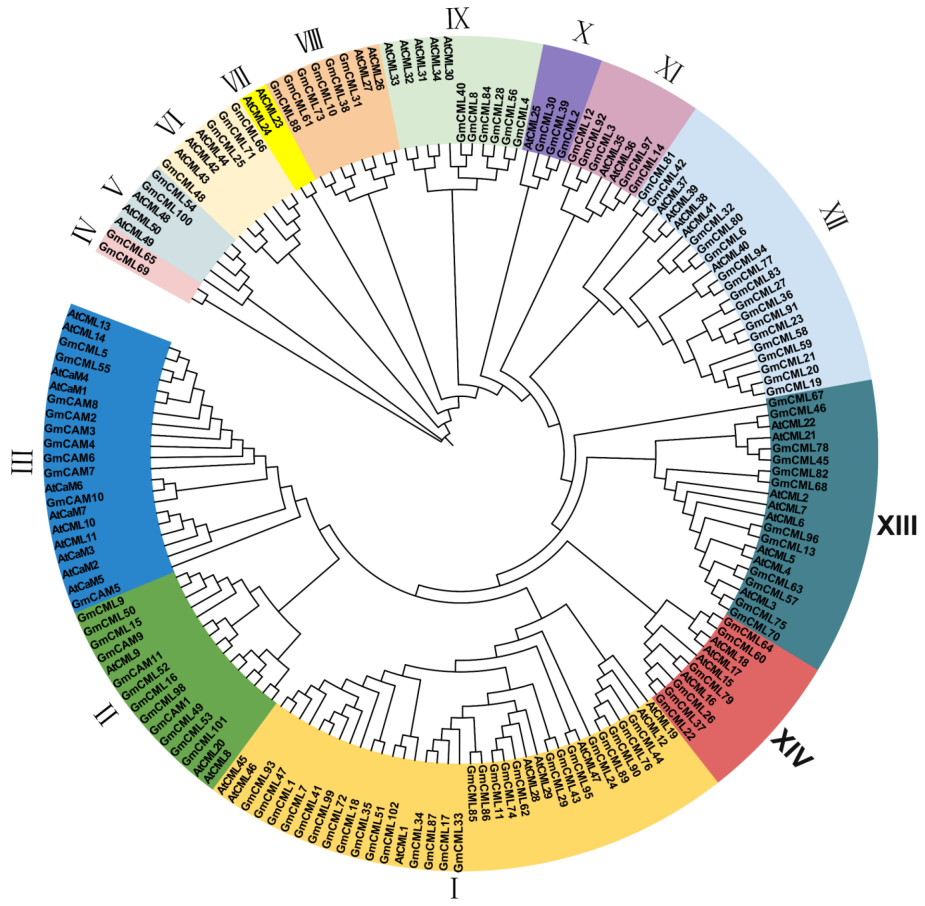


FIGURE 1
Phylogenetic tree of CAM/CML gene family members. Phylogenetic trees were built using MEGA-X software following the neighbor-joining (NJ) method and 1000 bootstrap replications were performed. Visualizing evolutionary trees using EvolView. All CAM/CML members are divided into 14 groups and presented in different colors.

containing only 2 members. Group IV exclusively comprises *GmCML* proteins (*GmCML*_65 and *GmCML*_69), while Group VII contained only *AtCML* proteins (*AtCML*_23 and *AtCML*_24). This species-specific clustering pattern suggests recent gene duplication events following the Arabidopsis-soybean divergence, potentially indicating lineage-specific functional adaptations. The phylogenetic distribution reveals several key evolutionary patterns. First, the extensive diversification of CML proteins (152 total) compared to CAM proteins (18 total) demonstrates significant sub functionalization following ancient duplication events. Second, the mixed species composition in most groups indicates that major gene family expansions occurred before monocot-dicot divergence. Third, the isolated positioning of the *GmCML* pair in Group IV, separated by long branch lengths from other clades, suggests potential neofunctionalization and acquisition of specialized calcium-sensing roles unique to legumes. Bootstrap values exceeding 85% for major nodes confirm the robustness of these phylogenetic relationships. The clear separation between CAM and CML lineages, combined with species-specific clustering patterns in smaller groups, provides a framework for understanding the evolutionary forces shaping calcium signaling diversity in flowering plants.

3.3 Conserved motif and gene structure analysis of CAM/CML genes

A comprehensive motif analysis of soybean CAM/CML genes was performed using the MEME suite to elucidate conserved structural and functional domains within this important calcium-binding protein family. Position weight matrix analysis revealed distinct amino acid conservation patterns within each motif (Figure 2a). The analysis identified five distinct conserved motifs across 15 CAM/CML genes from the Williams 82 reference genome (Wm82.gnm4.ann1), with statistical significance values ranging from 1.93 × 10^-40 to 1.27 × 10^-94 (Figure 2c). Motif distribution analysis revealed differential patterns of conservation across the gene family (Figure 2b). Motifs 1 and 2 demonstrated near-universal presence, occurring in 113 and 112 genes respectively, indicating their fundamental importance in CAM/CML protein structure and function. Motif 4 was present in 73 genes, while Motif 3 occurred in 60 genes. Notably, Motif 5 showed the most restricted distribution, appearing in only 24 genes, suggesting a specialized functional role within a subset of the CAM/CML family. Analysis of motif combinations revealed a modular architecture pattern across the CAM/CML gene family

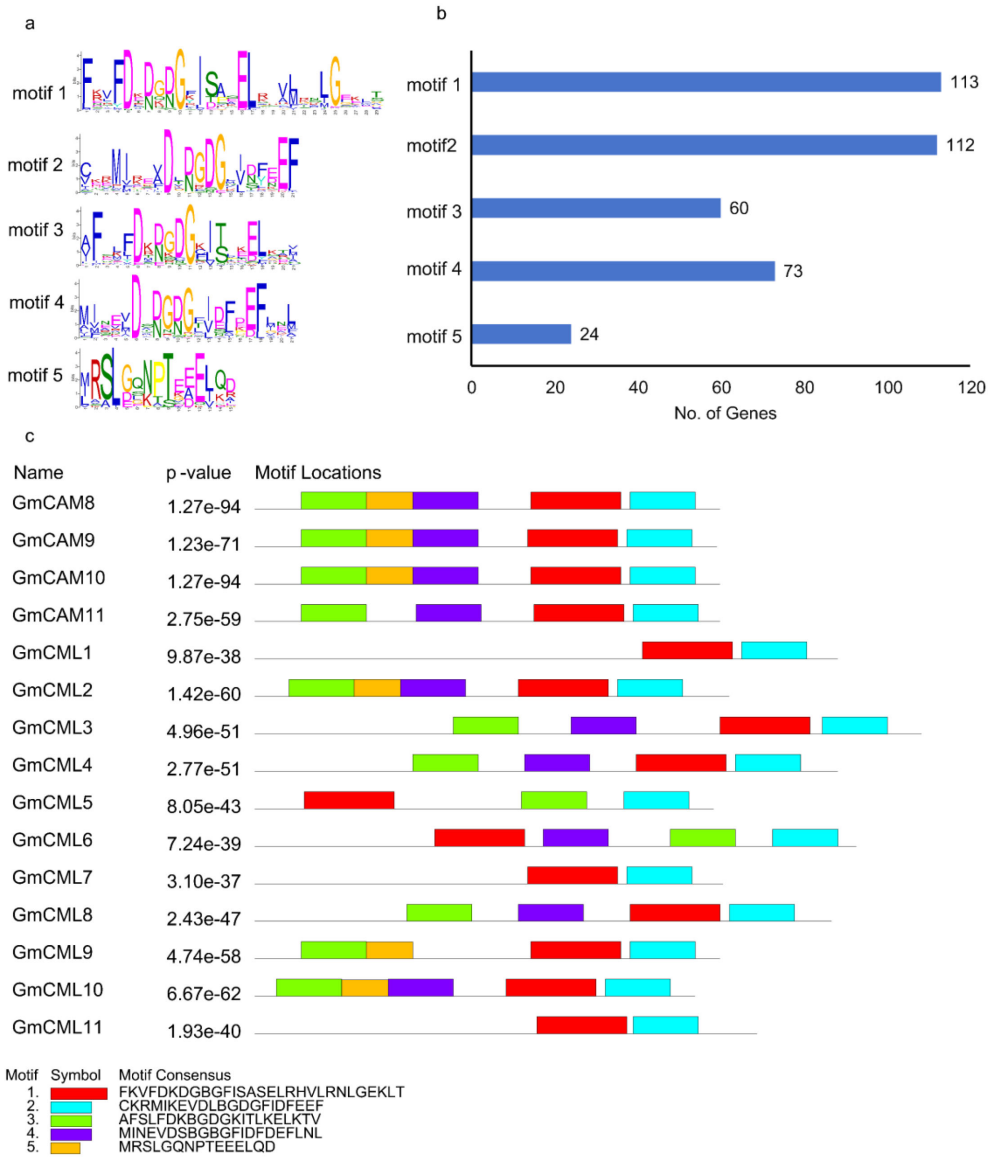


FIGURE 2
Conserved motif identification and structural organization of soybean CAM/CML proteins. **(a)** Position weight matrices showing amino acid conservation patterns for the five identified motifs. Letter height represents the relative frequency and information content at each position. **(b)** Distribution frequency of motifs across the CAM/CML gene family showing the number of genes containing each motif. **(c)** Motif architecture and genomic organization of 15 CAM/CML genes with statistical significance values (p-values). Colored blocks represent different motifs within each protein sequence.

(Figure 2c). Individual proteins contained between one and five motifs in various combinations, reflecting the evolutionary plasticity of this gene family. The most statistically significant proteins, including *GmCML 8*, *GmCML10* (both p = 1.27 × 10^-94), and *GmCML9* (p = 1.23 × 10^-71), displayed complex motif arrangements with multiple domain repeats. Several distinct architectural patterns emerged: (1) proteins containing primarily Motifs 1 and 2, representing the core CAM/CML structure; (2) proteins with additional Motifs 3 and/or 4, suggesting expanded functional capacity; and (3) a specialized subset containing Motif 5, potentially representing functionally divergent family members. This modular organization indicates that CAM/CML proteins have evolved through domain duplication and recombination events, allowing for functional diversification while maintaining essential calcium-binding capabilities.

The exceptionally high statistical significance of motif conservation (all p-values < 10^-30) demonstrates strong evolutionary pressure to maintain these structural elements, consistent with their critical role in calcium-mediated cellular processes. The universal presence of Motifs 1 and 2 across nearly all analyzed genes indicates these represent the essential functional core of CAM/CML proteins, likely corresponding to canonical EF-

hand calcium-binding domains. The restricted distribution of Motif 5 to a subset of genes suggests functional sub familiarization within the CAM/CML family. Proteins containing this motif may possess specialized roles in specific calcium signaling pathways or exhibit distinct regulatory mechanisms. The intermediate distribution patterns of Motifs 3 and 4 indicate additional layers of functional complexity, potentially conferring differential calcium sensitivity, protein-protein interaction capabilities, or subcellular localization signals. The identification of 15 CAM/CML genes with conserved motifs across multiple chromosomes indicates significant gene family expansion in soybean. The modular nature of motif combinations suggests that this expansion occurred through various mechanisms including tandem duplication, segmental duplication, and subsequent domain shuffling events.

3.4 Gene structure organization and domain architecture of CAM/CML family members

Gene structure analysis of the soybean CAM/CML family revealed distinct organizational patterns across 11 *GmCAM* and 102 *GmCML* genes (Figure 3). CAM genes exhibited relatively uniform structures with compact organization, while CML genes showed remarkable diversity, ranging from simple architectures to complex, multi-domain organizations extending over 8,000 base pairs. The five conserved motifs identified through MEME analysis were mapped to their genomic locations within gene structures. Motifs 1 and 2 were consistently located within coding sequences across all genes, confirming their essential functional roles. Motif 5 was predominantly found in CAM subfamily members, supporting functional specialization between CAM/CML subfamilies. Domain annotation revealed multiple EF-hand domain subtypes across the family, including *EFh_PEF*_Group_I, *EFh_SPARC_EC* superfamily, and *EFh_CREC* superfamily domains (Figure 3). *PTZ00184* and *PTZ00183* superfamily domains were identified in multiple members, suggesting additional functional capabilities beyond calcium binding. The FRQ1 domain was detected in a subset of genes, potentially conferring specialized regulatory functions. Most genes contained multiple EF-hand domains arranged in tandem, consistent with the typical architecture of calcium-binding proteins, which require multiple coordination sites. The modular arrangement supports evolution through domain duplication and recombination mechanisms. CAM genes demonstrate uniform structures with consistent domain arrangements, reflecting their conserved role as primary calcium sensors. The specific presence of Motif 5 in CAM genes provides a molecular basis for their unique regulatory properties. CML genes exhibited greater structural diversity with varying EF-hand domain combinations and accessory motifs, correlating with their functional specialization for distinct cellular processes. The structural diversity observed across the CAM/CML family demonstrates the evolutionary mechanisms that have shaped functional diversification, with CAM genes under stronger purifying selection while CML genes have undergone extensive structural innovation to generate the current functional versatility of plant calcium signaling networks.

3.5 Chromosomal distribution and gene structure analysis of *GmCAMS/CML*

A genome-wide analysis revealed the presence of 11 CAM and 102 CML genes distributed across all 20 chromosomes of the soybean genome. The distribution pattern exhibited significant variation, with chromosomes Gm11, Gm17, and Gm04 containing the highest number of CML genes, while the CAM genes were confined to only 7 chromosomes (Gm02, Gm03, Gm05, Gm13, Gm14, Gm19, and Gm20). Notably, we identified multiple clustering patterns of CML genes in specific chromosomal regions, suggesting potential tandem duplication events during evolution. The substantial expansion of the CML gene family (102 members) compared to the more conserved CAM family (11 members) suggests functional diversification following the whole-genome duplication events known to have occurred in soybean. This comprehensive chromosomal mapping provides important insights into the evolutionary history and potential functional specialization of calcium-signaling genes in Glycine max (Figure 4).

3.6 Cis-acting element analysis

Comprehensive analysis of the 2 kb upstream sequences of *GmCaM/CML* genes revealed a striking enrichment of hormone-responsive cis-acting elements (Figure 5a). Abscisic acid (ABA) responsive elements constituted the largest proportion (254 elements, 31.2%), followed closely by methyl jasmonate (MeJA) responsive elements (225 elements, 27.7%). Gibberellin-responsive elements (93 elements, 11.4%) and auxin-responsive elements (64 elements, 7.9%) were also well-represented. Elements associated with biotic and abiotic stress responses comprised a notable fraction, with defense/stress-responsive elements and salicylic acid-responsive elements accounting for 6.5% and 6.3%, respectively. *MYBHv1* binding sites (5.3%) and meristem expression elements (3.4%) were less abundant but present across multiple genes. Hierarchical clustering analysis revealed distinct cis-regulatory patterns among *GmCaM/CML* family members (Figure 5b), suggesting potential functional specialization. Several genes exhibited strong enrichment for specific hormone-responsive elements, while others displayed more diverse regulatory signatures. This regulatory diversity likely underpins the multifaceted roles of calmodulin and calmodulin-like proteins in calcium-mediated signaling networks, particularly in hormone signaling pathways and stress responses. The prevalence of hormone-responsive elements highlights the pivotal role of *GmCaM/CML* proteins as integration nodes between calcium signaling and hormone-regulated processes in soybean.

FIGURE 3
Gene structure organization and domain architecture of soybean CAM/CML family members. Schematic representation showing the complete gene structures of 11 *GmCAM* and 102 *GmCML* genes, including coding sequences (CDS), untranslated regions (UTR), and conserved protein domains. Motifs 1–5 identified through MEME analysis are mapped to their genomic locations. Protein domains include EF-hand domains (EFh) and their subtypes, PTZ superfamily domains, and FRQ1 domains. Scale bars indicate nucleotide positions. The organization demonstrates the modular architecture and evolutionary diversification of the CAM/CML gene family, with CAM genes showing more uniform structures while CML genes display remarkable structural diversity.

3.7 Extensive gene duplication and evolutionary conservation underpin *GmCaM/CML* gene family expansion

Collinearity analysis revealed substantial duplication events contributing to the expansion of the *CaM/CML* gene family in soybean (Figure 6a). We identified 117 pairs of homologous *GmCaM/CML* genes distributed across the soybean genome, visualized as connecting lines in the circular diagram. This extensive duplication pattern likely resulted from the two whole-genome duplication events in soybean's evolutionary history. Comparative genomic analysis between soybean and model plant species uncovered significant evolutionary relationships in the *CaM/CML* gene family (Figure 6b). We identified 67 pairs of homologous CaM/CML genes between soybean and Arabidopsis, and 32 pairs between soybean and rice. This differential homology pattern reflects the closer evolutionary relationship between soybean and Arabidopsis (both dicots) compared to rice (monocot). To assess evolutionary selection pressure, we calculated Ka/Ks ratios for all homologous gene pairs. Notably, all gene pairs exhibited Ka/Ks ratios below 1.0, indicating strong purifying selection (Supplementary Table S4). This evolutionary constraint suggests functional conservation of CaM/CML proteins across species, highlighting their fundamental importance in calcium-mediated signaling processes throughout plant evolution. The greater number of homologous relationships between soybean and Arabidopsis compared to rice aligns with established phylogenetic relationships between these species and suggests lineage-specific expansion and conservation patterns in the plant CaM/CML gene family.

3.8 Tissue-specific expression reveals functional diversification of *GmCaM/CML* genes

Analysis of RNA-Seq data across soybean tissues revealed distinct spatial expression patterns among *GmCaM/CML* family members (Supplementary Figure S1). Hierarchical clustering identified two major expression clusters: one predominantly

10.3389/fpls.2025.1633325

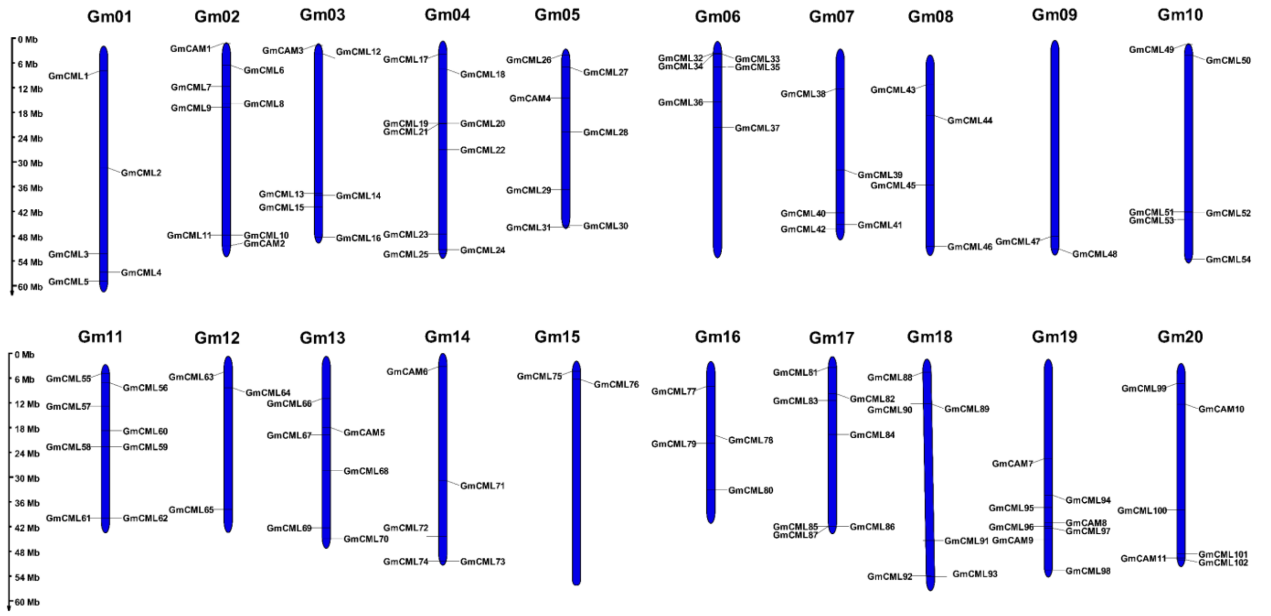


FIGURE 4
Chromosomal distribution of calmodulin (CAM) and calmodulin-like (CML) genes in the soybean genome. The chromosomal locations of 11 *GmCAM* and 102 *GmCML* genes were mapped across the 20 soybean chromosomes (Gm01-Gm20). Yellow bars represent chromosomes with the relative positions of genes indicated in megabases (Mb) from top (0 Mb) to bottom (60 Mb). Gene names are displayed adjacent to their respective positions. Note the uneven distribution pattern, with some chromosomes harboring multiple genes while others contain fewer. Clusters of genes in proximity suggest tandem duplication events, while the widespread distribution across all chromosomes reflects the importance of calcium signaling throughout the soybean genome.

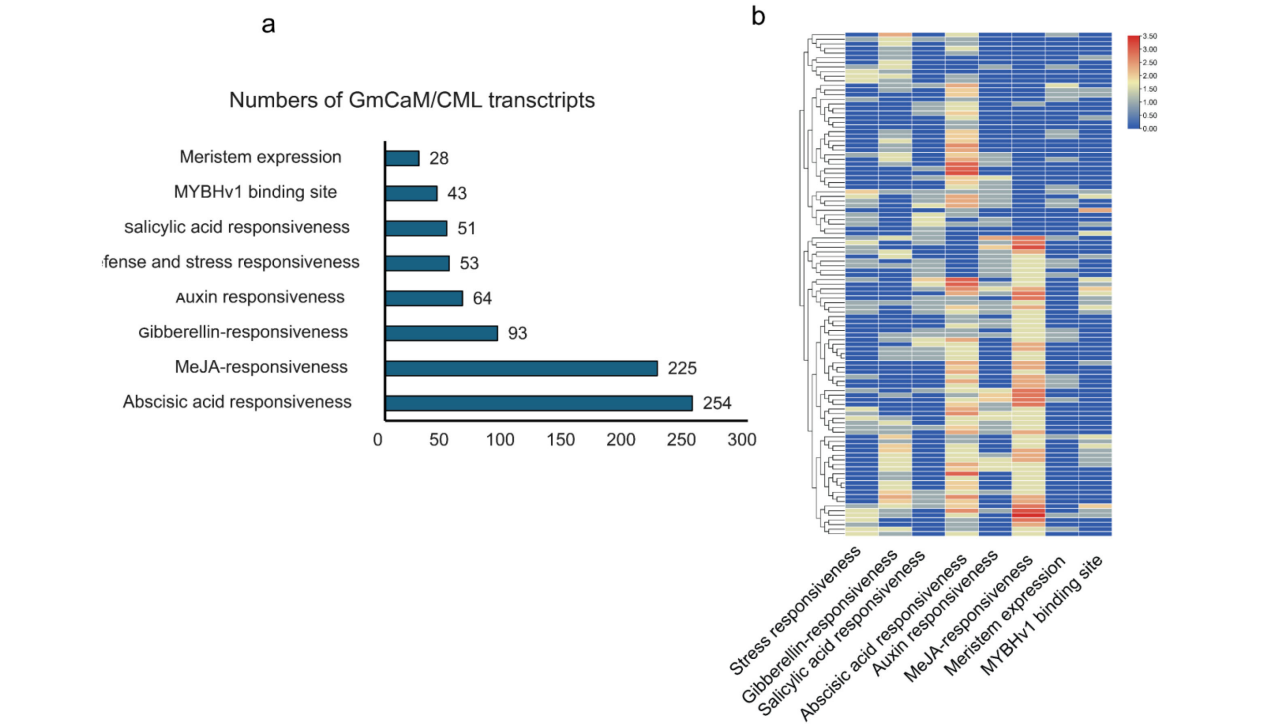


FIGURE 5
Cis-regulatory element analysis of *GmCaM/CML* genes. **(a)** Distribution of cis-acting elements in the 2 kb upstream regions of *GmCaM/CML* genes, categorized by function. **(b)** Hierarchical clustering heatmap showing the distribution pattern of different cis-regulatory elements across *GmCaM/CML* family members. Color intensity represents the abundance of each element type, with red indicating higher abundance.

10.3389/fpls.2025.1633325

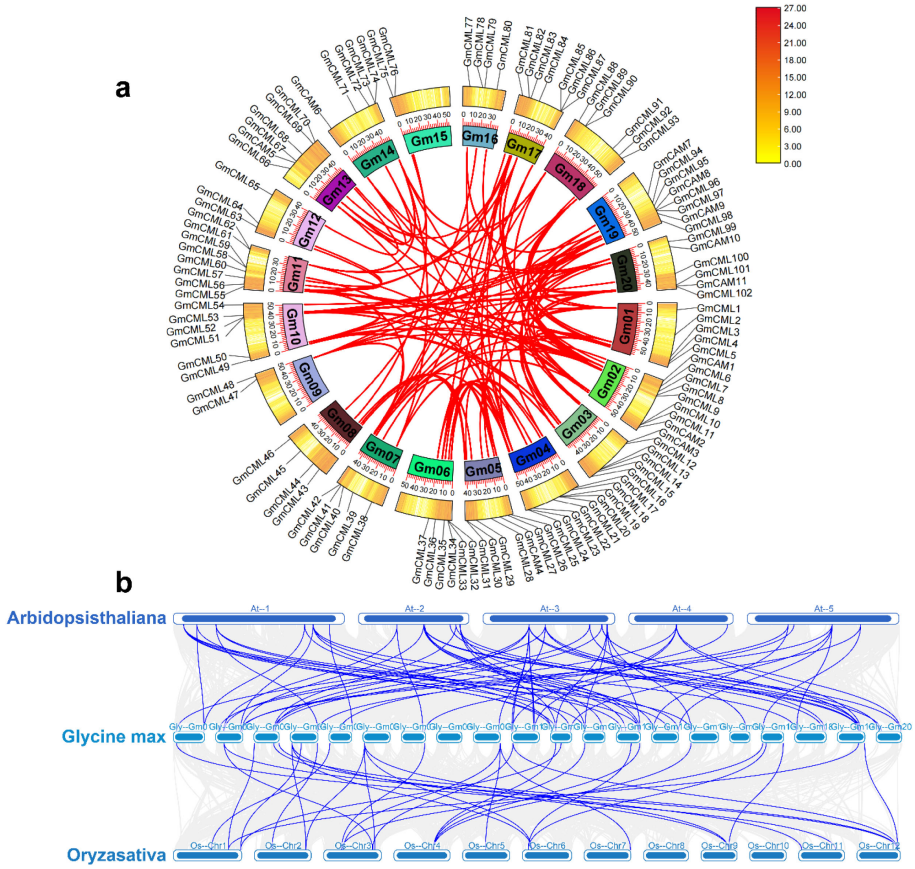


FIGURE 6
Collinearity analysis reveals duplication patterns and evolutionary conservation of the CaM/CML gene family. **(a)** Circular diagram showing homologous relationships (red lines) among 117 pairs of GmCaM/CML genes across soybean chromosomes. Colored blocks represent different chromosomes with gene distributions. Colored blocks represent different chromosomes with gene distributions. The color scale (0.00–27.00) indicates gene density per chromosomal region, where yellow represents low gene density and dark red represents high gene density. **(b)** Syntenic relationships of CaM/CML genes between Arabidopsis thaliana (top), Glycine max (middle), and Oryza sativa (bottom). Blue lines connect homologous genes across species, illustrating patterns of evolutionary conservation and divergence.

showing low expression across most tissues (upper cluster including *GmCML15* through *GmCAM9*), and another exhibiting moderate to high expression in various tissue combinations (lower cluster including *GmCML19* through *GmCML81*) (Figure 7). Several genes displayed striking tissue-specific expression profiles. *GmCML81* demonstrated robust expression across vegetative tissues but minimal activity in seeds. *GmCML83* showed high expression across multiple tissues, particularly in roots. Conversely, *GmCML22* exhibited preferential expression in flowers and roots, while *GmCML94* showed enhanced expression in leaves and flowers, suggesting specialized functions in these tissues. Notably, while phylogenetically related genes often shared similar expression patterns, we observed several instances of expression divergence among duplicated genes. For example, *GmCML83* maintained high expression across all tissues examined, whereas its homolog *GmCML94* was expressed at substantially lower levels. This expression divergence following gene duplication suggests neo- or sub-functionalization events contributing to the functional diversification of the *GmCaM/CML* gene family in soybean.

3.9 Differential expression profiles of 19 calmodulin-like genes in resistant and susceptible soybean varieties in response to *C. sojina* infection

Transcriptome analysis of 19 CML genes in Glycine max revealed distinct expression patterns correlated with resistance to *C. sojina* infection (Figure 8). Expression profiles clustered into three categories across the five resistant and five susceptible varieties: high-magnitude (GmCML32, GmCML65, GmCML77), moderate (GmCML7, GmCML15, GmCML66, GmCML83, GmCML94), and low-variable (GmCML19, GmCML22, GmCML23, GmCML41, GmCML58, GmCML69). GmCML77 exhibited the highest expression amplitude (10 relative units) in resistant varieties, particularly in samples 24JD210, 24JD697, and 24JD890. Interestingly, GmCML32 and GmCML65 exhibited substantial dynamic ranges (10 and 7.5 units, respectively), with significant upregulation in resistant varieties following infection. Conversely, *GmCML23* maintained a relatively stable expression (±

2 units) regardless of the resistance phenotype, suggesting a constitutive rather than an inducible role. Temporal expression patterns following *C. sojina* challenge fell into three categories: progressive increase (*GmCML15, GmCML32*), bell-shaped (*GmCAM9, GmCML27, GmCML30*), and oscillatory (*GmCML22, GmCML69, GmCML77*). The resistant varieties showed distinctive early upregulation of *GmCML94 and GmCML99*, while late-stage infection was characterized by elevated expression of *GmCML7, GmCML32*, and *GmCML65* in these same varieties (Figure 8). Five genes (*GmCML19*, *GmCML22, GmCML41, GmCML58, GmCML69*) demonstrated negative expression values in susceptible varieties relative to resistant ones, indicating potential suppression during pathogen infection. This was most pronounced in *GmCML22*, which exhibited the widest expression range (-5 to +5 units) between resistant and susceptible varieties. Notably, these repression patterns were primarily associated with early infection stages in susceptible varieties, suggesting pathogen-induced transcriptional interference. Sample-specific expression analysis revealed differential responses among resistant varieties (24JD210, 24JD697, and 24JD890) compared to susceptible varieties (24JD829, 24JD882, and 24JD892) (Figure 8). This divergence suggests functional specialization among CML family members may contribute to *C. sojina* resistance. Particularly, the coordinated expression of *GmCML77, GmCML81, and GmCML83* in resistant varieties points to a potential calcium-signaling cascade essential for mounting effective defense responses against this economically significant pathogen.

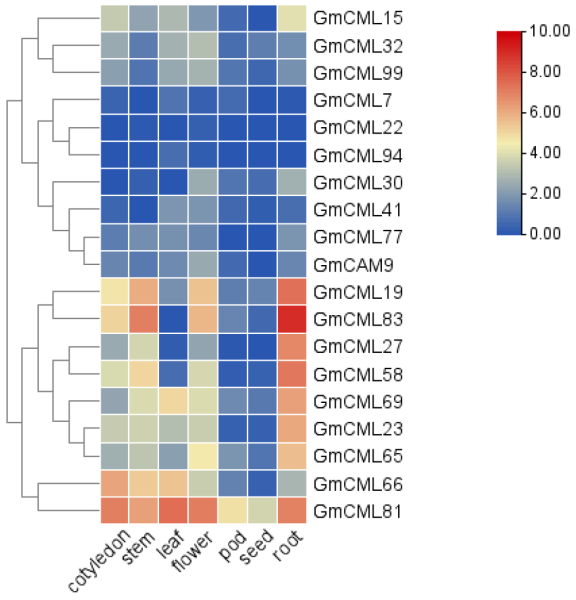


FIGURE 7
Tissue-specific expression profiles of GmCaM/CML genes. Heatmap showing expression patterns of selected GmCaM/CML genes across seven soybean tissues (cotyledon, stem, leaf, flower, pod, seed, and root) based on RNA-Seq data. Hierarchical clustering reveals groups with similar expression patterns. Color scale indicates normalized expression levels, with blue representing low expression and red representing high expression (scale 0-10).

4 Discussion

In this study, we conducted a genome-wide identification of *GmCaM* and GmCML genes in soybean, leading to the discovery of 11 *GmCAMs* and 102 *GmCMLs*, along with their collinearity, structures, chromosomal locations, and expression patterns across various tissues. Additionally, we analyzed the differential expression profiles of *GmCaM* and *GmCML* genes in response to SMV and *C. sojina* infection. This study offers comprehensive insights into soybean's *GmCaM* and *GmCML* gene families.

4.1 Evolutionary insights and family expansion of soybean CAM/CML genes

Our comprehensive analysis identified 113 CAM/CML genes (11 *GmCAMs* and 102 *GmCMLs*) in the soybean genome, representing one of the most prominent CAM/CML families reported in plants. This extensive gene family size compared to Arabidopsis (7 *AtCAMs* and 50 *AtCMLs*) (Mccormack et al., 2005), and rice (5 *OsCAMs* and 32 *OsCMLs*) (Boonburapong and Buaboocha, 2007), likely reflects the paleopolyploid nature of the soybean genome, which has undergone two whole-genome duplication events approximately 59 and 13 million years ago (Schmutz et al., 2010). The collinearity analysis revealed that most *GmCAM/CML* genes (117 pairs) resulted from these duplication events, significantly contributing to family expansion. The phylogenetic analysis revealed 14 distinct groups of varying sizes and compositions, indicating different gene retention rates and loss following duplication events. The presence of mixed species composition in most groups suggests that significant diversification of the CAM/CML family occurred before the divergence of monocots and dicots, estimated at 140–150 million years ago (Chaw et al., 2004; Liu et al., 2024). However, the species-specific clustering observed in some groups (e.g., Group IV, which contains only *GmCMLs*) suggests more recent lineage-specific duplication and potential neofunctionalization. The extremely low Ka/Ks ratios (<1.0) for all homologous gene pairs indicate strong purifying selection pressure, suggesting functional conservation despite extensive duplication. This is consistent with the essential role of calcium signaling in fundamental cellular processes and stress responses (Ranty et al., 2006). Interestingly, the Ka/Ks ratios were generally lower for *GmCAM* genes compared to *GmCML* genes, reflecting the higher evolutionary conservation of canonical calmodulins, which are likely to maintain core calcium-sensing functions.

4.2 Structural diversity and functional implications

The structural analysis revealed significant diversity in gene architecture and protein features across the GmCAM/CML family. While most *GmCAMs* (10/11) contain introns, a majority of

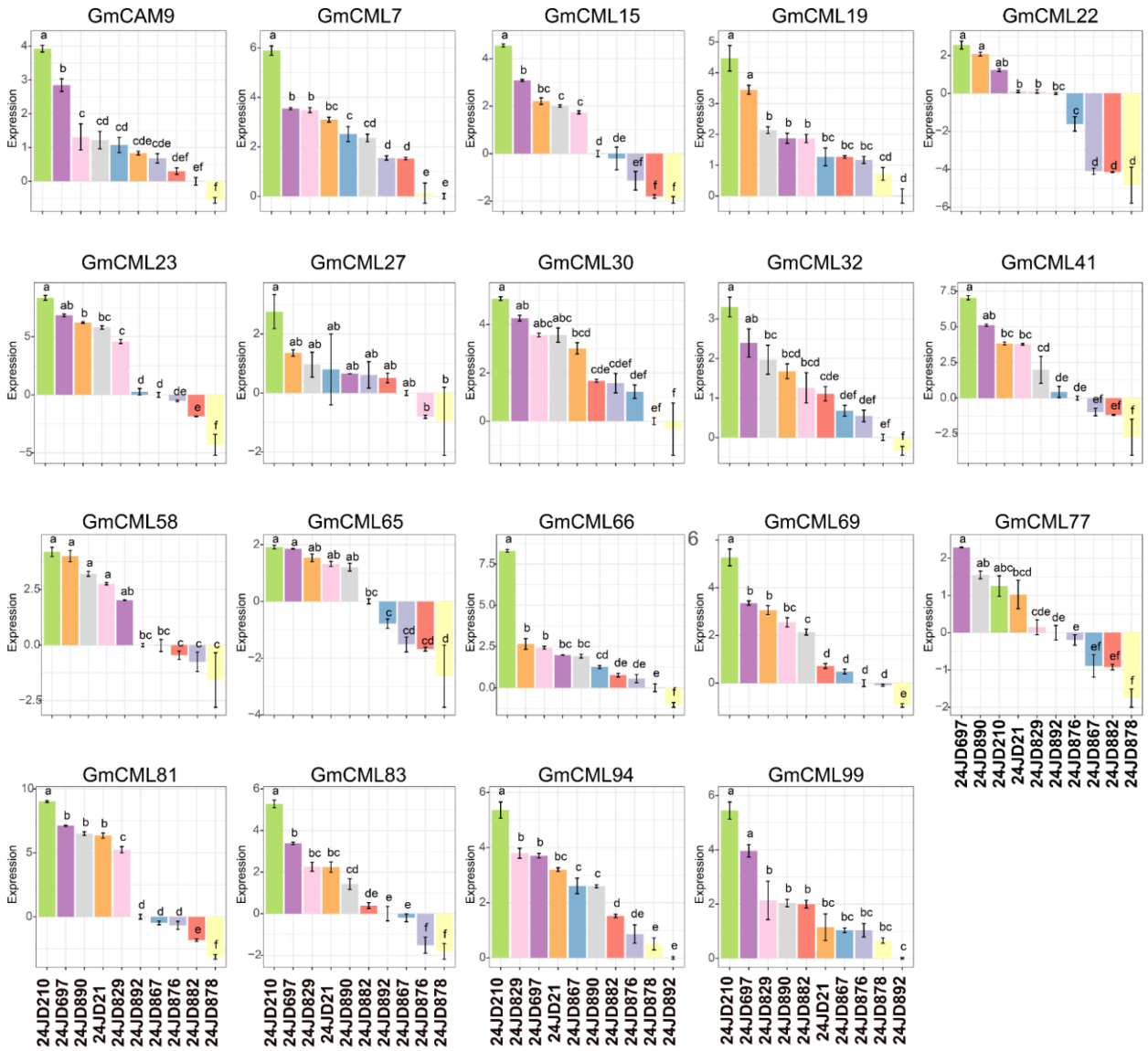


FIGURE 8
Differential expression profiles of 19 calmodulin-like genes in resistant and susceptible soybean varieties in response to *C. sojina* infection.

GmCMLs (71/102, 69.6%) are intron-less. This pattern is consistent with observations in Arabidopsis and rice (Mccormack et al., 2005; Boonburapong and Buaboocha, 2007), suggesting that the intron-less structure of most CMLs may represent an ancestral state or result from retro position events during evolution. The intron-less nature could potentially allow for more rapid expression in response to stress stimuli, as intron splicing can be rate-limiting during transcription (Jeffares et al., 2008). At the protein level, all identified *GmCAMs* contain four canonical EF-hand domains, consistent with their role as primary calcium sensors. In contrast, *GmCMLs* exhibit greater diversity in the number of EF-hand domains (1-6) and sequence, potentially reflecting functional diversification. The conservation of specific motifs (particularly Motif 1 present in all members and Motif 5 exclusive to *GmCAMs*) suggests distinct functional modules within these proteins. The structural divergence between CAMs and CMLs likely underlies their differential calcium-binding properties and target protein interactions, contributing to the specificity and versatility of calcium signaling responses (Gifford et al., 2007).

The striking structural divergence between *GmCAMs* and *GmCMLs* provides compelling evidence for neofunctionalization following gene duplication events in the soybean CAM/CML gene family (Yang et al., 2024). The contrasting intron-exon organization patterns, with 90.9% of *GmCAMs* retaining introns versus 69.6% of *GmCMLs* being intronless, suggest asymmetric evolutionary trajectories in which duplicated genes acquired distinct functional properties (Glick et al., 2024). The predominant intron-less structure of *GmCMLs* likely represents a key adaptive innovation that enabled specialization for rapid pathogen responses by eliminating time-consuming splicing processes during transcription (Johnson, 2019). This structural streamlining, coupled with diversification in EF-hand domain numbers (1–6 domains in CMLs vs. canonical 4 in CAMs)

and sequence divergence in calcium-binding regions, indicates coordinated neofunctionalization that enabled CML proteins to acquire specialized pathogen-sensing capabilities (Zhao et al., 2013). The evolutionary success of this neofunctionalization is demonstrated by the massive expansion of the CML subfamily (102 genes) compared to the conserved CAM subfamily (11 genes), suggesting that the acquisition of pathogen defense functions provided significant adaptive advantages. The identification of 15 *GmCML* genes responding to both viral and fungal pathogens further support the hypothesis that neofunctionalization enabled the evolution of broad-spectrum pathogen recognition capabilities, representing a classic example of how gene duplication followed by functional divergence can generate evolutionary novelty in plant defense systems.

4.3 Transcriptional regulation and stress responsiveness

The analysis of cis-acting elements in promoter regions revealed extensive enrichment of hormone-responsive elements, particularly those associated with abscisic acid (31.2%) and methyl jasmonate (27.7%) responses. This suggests that *GmCAM/CML* genes are integrated into hormone signaling networks mediating stress responses, consistent with previous reports in Arabidopsis and rice (Poovaiah et al., 2013). The significant presence of defense and stress-responsive elements (6.5%) and salicylic acid-responsive elements (6.3%) further supports their involvement in biotic stress responses. Interestingly, the dual pathogen-responsive genes identified in our study showed significant enrichment of specific cis-elements compared to the overall family, particularly W-box elements (WRKY binding sites) and TCA elements (salicylic acid-responsive). This enrichment suggests that these genes may be directly regulated by defense-related transcription factors such as WRKYs, which are known to play central roles in plant immunity (Pandey and Somssich, 2009; Wang et al., 2024). The co-occurrence of multiple hormone-responsive elements in these promoters indicates complex hormonal regulation, potentially involving crosstalk between salicylic acid, jasmonic acid, and abscisic acid pathways during pathogen defense.

4.4 CAM/CML genes in soybean dual resistance to SMV and *C. sojina*

A major finding of our study is the identification of 19 *GmCAM/CML* genes responsive to both SMV and *C. sojina* infection, suggesting their involvement in a common calcium-dependent defense mechanism. This is particularly noteworthy given the distinct nature of these pathogens—SMV is an RNA virus that propagates through the phloem, while *C. sojina* is a necrotrophic fungal pathogen that directly penetrates and kills host cells. The identification of calcium signaling components responsive to both pathogens suggest a role in broad-spectrum disease resistance. Among these dual-responsive genes, *GmCAM4*, *GmCML23*, and *GmCML47* showed the strongest correlation with resistance phenotypes across multiple varieties. *GmCAM4* has been previously implicated in soybean defense responses (Park et al., 2004), but our study is the first to demonstrate its role in resistance to both viral and fungal pathogens. The early and transient expression pattern of *GmCAM4* suggests its involvement in initial signal perception and transmission, consistent with the role of canonical calmodulins in rapid calcium signal decoding (Ranty et al., 2016). The consistent upregulation of *GmCML23* and *GmCML47* in resistant varieties in response to both pathogens, coupled with their strong correlation with resistance phenotypes, indicates their potential roles as key regulators in soybean immune responses. The sustained expression of *GmCML23* throughout the infection time course suggests its involvement in longer-term defense responses, possibly including transcriptional reprogramming and metabolic adjustments. In contrast, the intermediate expression pattern of *GmCML47* may reflect its role in the transition from early to late defense responses. The differential expression patterns of these genes between resistant and susceptible varieties could be attributed to several factors. First, resistant varieties may possess specific allelic variants of these genes that confer enhanced expression in response to pathogen infection. Second, regulatory networks controlling the expression of these genes may differ between resistant and susceptible varieties, possibly due to variations in transcription factor activity or chromatin accessibility. Third, the timing and magnitude of calcium signatures induced by pathogen perception may vary between resistant and susceptible varieties, affecting the downstream activation of CAM/CML genes. The potential regulatory relationship among *GmCAM4*, *GmCML23*, and *GmCML47*, as suggested by their sequential activation patterns, warrants further investigation. *GmCAM4*, being the earliest responder, might directly or indirectly regulate the expression of *GmCML47* and *GmCML23*. Alternatively, these genes might respond to different temporal aspects of calcium signatures induced during pathogen infection.

While we observed that several CAM/CML genes exhibited transcriptional responses to both pathogens, these findings should be interpreted with caution. The apparent similarities or differences in gene expression patterns between the two pathogen treatments cannot be quantitatively assessed due to the lack of standardized experimental controls and uniform methodological approaches. Therefore, any comparative conclusions regarding differential CAM/CML gene regulation in response to SMV versus *C. sojina* infection remain speculative and require validation through future studies employing consistent experimental designs and standardized conditions for both pathogen treatments.

5 Conclusion

This study provides the first comprehensive characterization of the CAM/CML gene family in soybean and its potential roles in disease resistance. We identified 113 CAM/CML genes (11 *GmCAMs* and 102 *GmCMLs*) and classified them into 14 distinct phylogenetic groups. Through an integrated analysis of gene

structure, protein motifs, chromosomal distribution, and expression patterns, we provide insights into the evolutionary history and functional diversification of this gene family. Our expression analyses in response to two major soybean pathogens, Soybean Mosaic Virus and *C. sojina*, revealed distinct expression patterns of CAM/CML genes associated with resistance. Notably, we identified 19 genes responsive to both pathogens, suggesting their involvement in a shared calcium-dependent defense mechanism. Among these, *GmCAM4*, *GmCML23*, and *GmCML47* showed the strongest correlation with resistance phenotypes, indicating their potential as key regulators of soybean immune responses. In conclusion, our findings enhance our understanding of calcium signaling in soybean disease resistance and identify promising targets for molecular breeding and genetic engineering approaches aimed at developing soybean varieties with enhanced resistance to multiple pathogens. While this study provides valuable insights into the structure and expression patterns of soybean CaM/CML genes, further functional studies including gene knockout, overexpression, and complementation analyses are needed to validate the specific roles of individual CaM/CML genes in pathogen resistance. Such functional validation will be crucial for translating these findings into practical breeding applications for developing disease-resistant soybean cultivars."

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

CZ: Conceptualization, Funding acquisition, Methodology, Validation, Writing – original draft, Writing – review & editing. YW: Data curation, Formal Analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing. RZ: Conceptualization, Formal Analysis, Investigation, Software, Validation, Writing – original draft, Writing – review & editing. RY: Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Visualization, Writing – original draft, Writing – review & editing. KZ: Conceptualization, Investigation, Methodology, Project administration, Validation, Writing – original draft, Writing – review & editing. XL: Conceptualization, Formal Analysis, Methodology, Software, Validation, Writing – original draft, Writing – review & editing. XW: Investigation, Methodology, Project administration, Resources, Writing – original draft, Writing – review & editing. FZ: Conceptualization, Formal Analysis, Funding acquisition, Supervision, Writing – original draft, Writing – review & editing. SFL: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Resources, Software, Supervision, Validation, Writing – original draft, Writing – review & editing. BZ: Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing. HR: Conceptualization, Formal Analysis, Funding acquisition, Project administration, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported by the Project funded by Agricultural Science and Technology Innovation Leaping Project in Heilongjiang Province (Grant No.CX25JC48); Natural Science Foundation of Heilongjiang Province of China (PL2024C027); Scientific Research Business Expenses of Heilongjiang Scientific Research Institutes (Grant No.CZKYF2025-1-C011).

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.2025.1633325/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Heatmap showing expression patterns of all identified CaM/CML genes.

References

Arfaoui, A., El Hadrami, A., Adam, L. R., and Daayf, F. (2016). Pre-treatment with calcium enhanced defense-related genes' expression in the soybean's isoflavones pathway in response to Sclerotinia sclerotiorum. *Physiol. Mol. Plant Pathol.* 93, 12–21. doi: 10.1016/j.pmpp.2015.11.004

Barro, J. P., Neves, D. L., Del Ponte, E. M., and Bradley, C. A. (2023). Frogeye leaf spot caused by Cercospora sojina: A review. *Trop. Plant Pathol.* 48, 363–374. doi: 10.1007/s40858-023-00583-8

Boonburapong, B., and Buaboocha, T. (2007). Genome-wide identification and analyses of the rice calmodulin and related potential calcium sensor proteins. *BMC Plant Biol.* 7, 1–17. doi: 10.1186/1471-2229-7-4

Boudsocq, M., and Sheen, J. (2013). CDPKs in immune and stress signaling. *Trends Plant Sci.* 18, 30–40. doi: 10.1016/j.tplants.2012.08.008

Chaw, S.-M., Chang, C.-C., Chen, H.-L., and Li, W.-H. (2004). Dating the monocot–dicot divergence and the origin of core eudicots using whole chloroplast genomes. *J. Mol. Evol.* 58, 424–441. doi: 10.1007/s00239-003-2564-9

Chen, C., Wu, Y., Li, J., Wang, X., Zeng, Z., Xu, J., et al. (2023). TBtools-II: A "one for all, all for one" bioinformatics platform for biological big-data mining. *Mol. Plant* 16, 1733–1742. doi: 10.1016/j.molp.2023.09.010

Cheng, H.-Q., Han, L.-B., Yang, C.-L., Wu, X.-M., Zhong, N.-Q., Wu, J.-H., et al. (2016). The cotton MYB108 forms a positive feedback regulation loop with CML11 and participates in the defense response against Verticillium dahliae infection. *J. Exp. Bot.* 67, 1935–1950. doi: 10.1093/jxb/erw016

Cheng, S.-H., Willmann, M. R., Chen, H.-C., and Sheen, J. (2002). Calcium signaling through protein kinases. The Arabidopsis calcium-dependent protein kinase gene family. *Plant Physiol.* 129, 469–485. doi: 10.1104/pp.005645

Choi, B. K., Koo, J. M., Ahn, H., Yum, H., Choi, C. W., Ryu, K. H., et al. (2005). Emergence of Rsv-resistance breaking Soybean mosaic virus isolates from Korean soybean cultivars. *Virus Res.* 112, 42–51. doi: 10.1016/j.virusres.2005.03.020

Day, I. S., Reddy, V. S., Shad Ali, G., and Reddy, A. (2002). Analysis of EF-hand-containing proteins in arabidopsis. *Genome Biol.* 3, 1–24. doi: 10.1186/gb-2002-3-10-research0056

Gao, X., Cox, K. L. Jr., and He, P. (2014). Functions of calcium-dependent protein kinases in plant innate immunity. *Plants* 3, 160–176. doi: 10.3390/plants3010160

Gifford, J. L., Walsh, M. P., and Vogel, H. J. (2007). Structures and metal-ion-binding properties of the Ca2+-binding helix–loop–helix EF-hand motifs. *Biochem. J.* 405, 199–221. doi: 10.1042/BJ20070255

Glick, L., Castiglione, S., Loewenthal, G., Raia, P., Pupko, T., and Mayrose, I. (2024). Phylogenetic analysis of 590 species reveals distinct evolutionary patterns of intron–exon gene structures across eukaryotic lineages. *Mol. Biol. Evol.* 41, msae248. doi: 10.1093/molbev/msae248

He, X., Liu, W., Li, W., Liu, Y., Wang, W., Xie, P., et al. (2020). Genome-wide identification and expression analysis of CaM/CML genes in Brassica napus under abiotic stress. *J. Plant Physiol.* 255, 153251. doi: 10.1016/j.jplph.2020.153251

Jeena, H., Singh, K. P., Rakhonde, G., Dev, M., Surbhi, K., and Aravind, T. (2024). "Diagnosis and management strategies for soybean diseases," in *Diseases of Field Crops: Diagnostics and Management* (Berlin/Heidelberg, Germany: Springer), 223–264.

Jeffares, D. C., Penkett, C. J., and Bähler, J. (2008). Rapidly regulated genes are intron poor. *Trends Genet.* 24, 375–378. doi: 10.1016/j.tig.2008.05.006

Johnson, N. T. (2019). *Leveraging Omics Data to Expand the Value and Understanding of Alternative Splicing* (Worcester Polytechnic Institute).

Kaur, A., Sharma, A., Verma, P. C., and Upadhyay, S. K. (2022). EF-hand domain-containing proteins in Triticum aestivum: Insight into their roles in stress response and signalling. *South Afr. J. Bot.* 149, 663–681. doi: 10.1016/j.sajb.2022.06.059

Kim, H. J., Kim, M.-J., Pak, J. H., Jung, H. W., Choi, H. K., Lee, Y.-H., et al. (2013). Characterization of SMV resistance of soybean produced by genetic transformation of SMV-CP gene in RNAi. *Plant Biotechnol. Rep.* 7, 425–433. doi: 10.1007/s11816-013-0279-y

Li, H., Liu, J., Yuan, X., Chen, X., and Cui, X. (2023). Comparative transcriptome analysis reveals key pathways and regulatory networks in early resistance of Glycine max to soybean mosaic virus. *Front. Microbiol.* 14, 1241076. doi: 10.3389/fmicb.2023.1241076

Li, Y., Liu, Y., Jin, L., and Peng, R. (2022). Crosstalk between Ca2+ and other regulators assists plants in responding to abiotic stress. *Plants* 11, 1351.

Liu, X., Zhang, C., Lamlom, S. F., Zhao, K., Abdelghany, A. M., Wang, X., et al. (2024). Genetic adaptations of soybean to cold stress reveal key insights through transcriptomic analysis. *Biology* 13, 856. doi: 10.3390/biology13110856

Luan, S. (2009). The CBL–CIPK network in plant calcium signaling. *Trends Plant Sci.* 14, 37–42. doi: 10.1016/j.tplants.2008.10.005

Ma, W., Smigel, A., Tsai, Y.-C., Braam, J., and Berkowitz, G. A. (2008). Innate immunity signaling: cytosolic Ca2+ elevation is linked to downstream nitric oxide generation through the action of calmodulin or a calmodulin-like protein. *Plant Physiol.* 148, 818–828. doi: 10.1104/pp.108.125104

Madhusudhan, B (2022). Effect of mosaic and mottle disease on yield and seed quality parameters of soybean.

Majidian, P., Ghorbani, H. R., and Farajpour, M. (2024). Achieving agricultural sustainability through soybean production in Iran: Potential and challenges. *Heliyon* 10 (4). doi: 10.1016/j.heliyon.2024.e26389

Mccormack, E., and Braam, J. (2003). Calmodulins and related potential calcium sensors of Arabidopsis. *New Phytol.* 159, 585–598. doi: 10.1046/j.1469-8137.2003.00845.x

Mccormack, E., Tsai, Y.-C., and Braam, J. (2005). Handling calcium signaling: arabidopsis CaMs and CMLs. *Trends Plant Sci.* 10, 383–389. doi: 10.1016/j.tplants.2005.07.001

Mohanta, T. K., Kumar, P., and Bae, H. (2017). Genomics and evolutionary aspect of calcium signaling event in calmodulin and calmodulin-like proteins in plants. *BMC Plant Biol.* 17, 1–19. doi: 10.1186/s12870-017-0989-3

Negi, N. P., Prakash, G., Narwal, P., Panwar, R., Kumar, D., Chaudhry, B., et al. (2023). The calcium connection: exploring the intricacies of calcium signaling in plant-microbe interactions. *Front. Plant Sci.* 14, 1248648. doi: 10.3389/fpls.2023.1248648

Pandey, S. P., and Somssich, I. E. (2009). The role of WRKY transcription factors in plant immunity. *Plant Physiol.* 150, 1648–1655. doi: 10.1104/pp.109.138990

Park, C. Y., Do Heo, W., Yoo, J. H., Lee, J. H., Kim, M. C., Chun, H. J., et al. (2004). Pathogenesis-related gene expression by specific calmodulin isoforms is dependent on NIM1, a key regulator of systemic acquired resistance. *Mol. Cells* 18, 207–213. doi: 10.1016/S1016-8478(23)13103-7

Park, H. C., Chun, H. J., Kim, M. C., Lee, S. W., and Chung, W. S. (2020). Identification of disease resistance to soft rot in transgenic potato plants that overexpress the soybean calmodulin-4 gene (GmCaM-4). *J. Plant Biotechnol.* 47, 157–163. doi: 10.5010/JPB.2020.47.2.157

Perochon, A., Aldon, D., Galaud, J.-P., and Ranty, B. (2011). Calmodulin and calmodulin-like proteins in plant calcium signaling. *Biochimie* 93, 2048–2053. doi: 10.1016/j.biochi.2011.07.012

Poovaiah, B., Du, L., Wang, H., and Yang, T. (2013). Recent advances in calcium/calmodulin-mediated signaling with an emphasis on plant-microbe interactions. *Plant Physiol.* 163, 531–542. doi: 10.1104/pp.113.220780

Ramlal, A., Harika, A., Jayasri, V., Subramaniam, S., Mallikarjuna, B. P., Raju, D., et al. (2024). Calmodulin: Coping with biotic and abiotic stresses in soybean (Glycine max (L.) Merr.). *Plant Stress* 14, 100602. doi: 10.1016/j.stress.2024.100602

Ranty, B., Aldon, D., Cotelle, V., Galaud, J.-P., Thuleau, P., and Mazars, C. (2016). Calcium sensors as key hubs in plant responses to biotic and abiotic stresses. *Front. Plant Sci.* 7, 327. doi: 10.3389/fpls.2016.00327

Ranty, B., Aldon, D., and Galaud, J.-P. (2006). Plant calmodulins and calmodulin-related proteins: multifaceted relays to decode calcium signals. *Plant Signaling Behav.* 1, 96–104. doi: 10.4161/psb.1.3.2998

Ren, H., Zhang, B., Zhang, F., Liu, X., Wang, X., Zhang, C., et al. (2024). Integration of physiological and transcriptomic approaches in investigating salt-alkali stress resilience in soybean. *Plant Stress* 11, 100375. doi: 10.1016/j.stress.2024.100375

Ren, H., Zhang, F., Zhu, X., Lamlom, S. F., Zhao, K., Zhang, B., et al. (2023). Manipulating rhizosphere microorganisms to improve crop yield in saline-alkali soil: a study on soybean growth and development. *Front. Microbiol.* 14, 1233351. doi: 10.3389/fmicb.2023.1233351

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

Tong, T., Li, Q., Jiang, W., Chen, G., Xue, D., Deng, F., et al. (2021). Molecular evolution of calcium signaling and transport in plant adaptation to abiotic stress. *International J. Mol. Sci.* 22, 12308.

Tuteja, N., and Mahajan, S. (2007). Calcium signaling network in plants: an overview. *Plant Signaling Behav.* 2, 79–85. doi: 10.4161/psb.2.2.4176

Vidhyasekaran, P., and Vidhyasekaran, P. (2020). Manipulation of calcium ion influx —Mediated immune signaling systems for crop disease management. *Plant Innate Immun. Signals Signaling Syst.: Bioeng. Mol. Manipulation Crop Dis. Manage.*, 23–49. doi: 10.1007/978-94-024-1940-5

Wang, J.-P., Munyampundu, J.-P., Xu, Y.-P., and Cai, X.-Z. (2015). Phylogeny of plant calcium and calmodulin-dependent protein kinases (CCaMKs) and functional analyses of tomato CCaMK in disease resistance. *Front. Plant Sci.* 6, 1075. doi: 10.3389/fpls.2015.01075

Wang, X., Zhang, C., Yuan, R., Liu, X., Zhang, F., Zhao, K., et al. (2024). Transcriptome profiling uncovers differentially expressed genes linked to nutritional quality in vegetable soybean. *PloS One* 19, e0313632. doi: 10.1371/journal.pone.0313632

Yang, D., Chen, T., Wu, Y., Tang, H., Yu, J., Dai, X., et al. (2024). Genome-wide analysis of the peanut CaM/CML gene family reveals that the AhCML69 gene is associated with resistance to Ralstonia solanacearum. *BMC Genomics* 25, 200. doi: 10.1186/s12864-024-10108-5

Yang, H., and Luo, P. (2021). Changes in photosynthesis could provide important insight into the interaction between wheat and fungal pathogens. *Int. J. Mol. Sci.* 22, 8865. doi: 10.3390/ijms22168865

Yang, T., and Poovaiah, B. (2003). Calcium/calmodulin-mediated signal network in plants. *Trends Plant Sci.* 8, 505–512. doi: 10.1016/j.tplants.2003.09.004

Zeng, H., Xu, L., Singh, A., Wang, H., Du, L., and Poovaiah, B. (2015). Involvement of calmodulin and calmodulin-like proteins in plant responses to abiotic stresses. *Front. Plant Sci.* 6, 600. doi: 10.3389/fpls.2015.00600

Zeng, H., Zhang, Y., Zhang, X., Pi, E., and Zhu, Y. (2017). Analysis of EF-hand proteins in soybean genome suggests their potential roles in environmental and nutritional stress signaling. *Front. Plant Sci.* 8, 877. doi: 10.3389/fpls.2017.00877

Zhang, C., Hajimorad, M., Eggenberger, A. L., Tsang, S., Whitham, S. A., and Hill, J. H. (2009a). Cytoplasmic inclusion cistron of Soybean mosaic virus serves as a virulence determinant on Rsv3-genotype soybean and a symptom determinant. *Virology* 391, 240–248. doi: 10.1016/j.virol.2009.06.020

Zhang, W., Zhou, R.-G., Gao, Y.-J., Zheng, S.-Z., Xu, P., Zhang, S.-Q., et al. (2009b). Molecular and genetic evidence for the key role of AtCaM3 in heat-shock signal transduction in Arabidopsis. *Plant Physiol.* 149, 1773–1784. doi: 10.1104/pp.108.133744

Zhao, Y., Liu, W., Xu, Y.-P., Cao, J.-Y., Braam, J., and Cai, X.-Z. (2013). Genome-wide identification and functional analyses of calmodulin genes in Solanaceous species. *BMC Plant Biol.* 13, 1–15. doi: 10.1186/1471-2229-13-70

Zhu, X., Perez, M., Aldon, D., and Galaud, J.-P. (2017). Respective contribution of CML8 and CML9, two arabidopsis calmodulin-like proteins, to plant stress responses. *Plant Signaling Behav.* 12, e1322246. doi: 10.1080/15592324.2017.1322246

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 29 August 2025
DOI 10.3389/fpls.2025.1630435

OPEN ACCESS

EDITED BY
Santosh Gudi,
North Dakota State University, United States

REVIEWED BY
Ajay Kumar Thakur,
Central Potato Research Institute (ICAR), India
Xinyi Wu,
Zhejiang Academy of Agricultural Sciences, China
Raghav Kataria,
Utah State University, United States

*CORRESPONDENCE
Hari Chand Sharma
hcsh19@gmail.com
Manish K. Pandey
Manish.Pandey@icrisat.org

RECEIVED 20 May 2025
ACCEPTED 29 July 2025
PUBLISHED 29 August 2025

CITATION
Moghiya A, Munghate RS, Sharma V, Mishra SP, Jaba J, Gaurav SS, Gangurde SS, Dube N, Rangari SK, Roychowdhury R, Gangashetty P, Sharma HC and Pandey MK (2025) Dissecting genomic regions and candidate genes for pod borer resistance and component traits in pigeonpea minicore collection.
Front. Plant Sci. 16:1630435.
doi: 10.3389/fpls.2025.1630435

Dissecting genomic regions and candidate genes for pod borer resistance and component traits in pigeonpea minicore collection

Abhinav Moghiya[1,2], R.S. Munghate[1], Vinay Sharma[1,2], Suraj Prashad Mishra[1], Jagdish Jaba[1], Shailendra Singh Gaurav[2], Sunil S. Gangurde[1], Namita Dube[1], Sagar Krushnaji Rangari[1], Rajib Roychowdhury[1], Prakash Gangashetty[1], Hari Chand Sharma[1]* and Manish K. Pandey[1]*

[1]Center of Excellence in Genomics and Systems Biology (CEGSB), and Center for Pre-Breeding Research (CPBR), International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Hyderabad, India, [2]Department of Genetics and Plant Breeding, Chaudhary Charan Singh University (CCSU), Meerut, India

Background: Pigeonpea is an important leguminous food crop primarily grown in tropical and subtropical regions of the world and is a rich source of high-quality protein. Biotic (weed, disease, and insect pests) and abiotic stresses have significantly reduced the production and productivity of pigeonpea. *Helicoverpa armigera*, also known as the pod borer, is a major pest in pigeonpea. A substantial investigation is needed to comprehend the genetic and genomic underpinnings of resistance to *H. armigera*. Genetic improvement by genomics-assisted breeding (GAB) is an effective approach for developing high-yielding *H. armigera*-resistant cultivars. Still, no genetic markers and genes linked to this key trait have been detected in pigeonpea. In this context, a set of 146 pigeonpea minicore accessions were evaluated for four *H. armigera*-resistant component traits, namely, pod borer resistance (PBR), days to 50% flowering (DF), days to maturity (DM), and grain yield (GY), for three consecutive seasons under field conditions.

Results: Phenotypic data of pod borer resistance and component traits, along with the whole-genome resequencing (WGRS) data for 4,99,980 single nucleotide polymorphisms (SNPs), were utilised to perform multi-locus genome-wide association study (GWAS) analysis. Two models [settlement of MLM under progressively exclusive relationship (SUPER) and fixed and random model circulating probability unification (FarmCPU)] detected 14 significant marker–trait associations (MTAs) for PBR and three component traits. The MTAs with significant effect were mainly identified on chromosomes CcLG02, CcLG04, CcLG05, CcLG07, and CcLG11. These MTAs were subsequently delineated with key candidate genes associated with pod borer resistance (*probable carboxylesterase 15*, *microtubule-associated protein 5*, *FAR1-RELATED SEQUENCE*, and *omega-hydroxypalmitate O-feruloyl transferase 4*), days to maturity (*RING-H2 finger protein ATL7* and *leucine-rich repeat receptor-like protein kinase*), and grain yield (*secretory carrier-associated membrane protein* and *glutaredoxin-C5 chloroplastic*).

Conclusion: These research findings reported significant MTAs and candidate genes associated with pod borer resistance and component traits. Further lab-

based pod bioassay screening identified four minicore accessions, namely, ICP 10503, ICP 655, ICP 9691, and ICP 9655 (moderately resistant genotypes), showing the least damage rating and larval weight gain %, compared to the susceptible checks. After validating the significant MTAs, the associated SNP markers can be effectively utilised in indirect selection, which offers potential gains for such quantitative traits with low heritability and can improve insect management more sustainably. The significant MTAs, candidate genes, and resistant accessions reported in this study may be utilised for the development of pod borer-resistant pigeonpea varieties.

KEYWORDS

marker-trait association, candidate gene discovery, genomic regions, mini-core collection, insect damage score, insect resistance

1 Introduction

Pigeonpea [*Cajanus cajan* (L.) Millsp.] is an important food legume crop in the arid and semi-arid regions of Asia and Africa. It is grown on 5.7 million hectares worldwide, with a production of 4.9 million tons (FAO, 2024). India, along with Malawi, Tanzania, Kenya, Uganda, and Myanmar, is a leading producer, contributing 78% of the global pigeonpea production. As one of the five major edible legumes, pigeonpea is used for edible purposes, animal feed, and firewood. It is an important source of protein, often used to supplement cereal-based diets (Kinhoégbè et al., 2022). Climate change presents a substantial risk to worldwide pigeonpea production, impacting both its nutritional quality and its ability to withstand various abiotic and biotic stresses. In India, pulses are vulnerable to approximately 150 insect pest species (Seetharamu et al., 2020), and globally, approximately 38 species of Lepidopteran insects harm pigeonpea (Shanower et al., 1999). Among the most damaging biotic stresses is the pod borer, *Helicoverpa armigera*, which severely affects crop growth and yield (Ghosh et al., 2017). Although pesticides can control the pod borer complex (PBC), the excessive use of chemical insecticides has resulted in insect resistance, secondary pest outbreaks, detrimental impacts on biodiversity, and negative environmental effects (Ambidi et al., 2021; Jaba et al., 2023). Therefore, developing pigeonpea varieties that are resistant to *H. armigera* is seen as the most effective solution to reduce pesticide use. Despite extensive screening of various pigeonpea genetic resources across Asia and Africa, no strong resistance against pod borer has been reported (Kambrekar, 2016). However, partial resistance has been reported in some cultivated genotypes, which have been utilised in pigeonpea breeding programs. While wild pigeonpea species confer higher pod borer resistance (PBR) compared to cultivated sources, transferring these resistance genes to cultivated varieties is limited to only a few wild species due to cross-incompatibility (Sharma, 2016; Singh et al., 2020). In earlier investigations, the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) minicore collection was screened, showing moderate resistance levels to the pod borer (Sharma et al., 2025, unpublished). These data have now been utilised for conducting a genome-wide association study (GWAS) and facilitating gene discovery.

The development and use of genomic tools can facilitate the selection of genotypes/breeding lines that are resistant to *H. armigera* using marker-assisted selection (MAS). However, there seems to be a lack of effort in identifying candidate genes and markers. Molecular markers are important for facilitating the transfer of insect-resistant genes to elite backgrounds, elucidating gene action, and minimising the negative effects of integrating undesirable genes from wild relatives due to linkage drag. Molecular breeding holds the potential to pyramid various sources of resistance that may not be efficiently selected by conventional breeding strategies due to phenotypic similarities, which can increase resistance levels and potentially develop resistant varieties (Sharma and Crouch, 2004). Recent breakthroughs in pigeonpea genomics research have resulted in the development of draft and telomere-to-telomere reference genomes (Varshney et al., 2012; Garg et al., 2022; Liu et al., 2024). Additionally, the accessibility of whole-genome sequencing (WGS) data (Varshney et al., 2017) and high-density Axiom *Cajanus* SNP arrays with 56K SNPs (Saxena et al., 2017) has significantly advanced genetic diversity, quantitative trait locus sequencing (QTL-seq), and genome-wide association study analysis. GWAS or association mapping has emerged as an important tool for identifying marker–trait associations (MTAs), candidate genes, and associated markers (Gudi et al., 2024; Sharma et al., 2024). Whole-genome resequencing (WGRS)-based GWAS is effective for identifying associated genomic regions and candidate genes related to specific traits in various legume species, including pigeonpea (Varshney et al., 2012; Xu et al., 2017; Kang et al., 2019). Recent studies have detected MTAs for flowering time (Kumar et al., 2022) and antioxidant properties (Megha et al., 2024). Similarly, meta-QTLs (MQTLs) were identified for agronomic traits, fertility restoration, disease resistance, and seed quality

traits (Halladakeri et al., 2023). This investigation utilised multi-season phenotyping data generated on diverse minicore accessions to identify significant MTAs and candidate genes linked with pod borer resistance. We highlighted the importance of using various resistance sources against pod borer damage, emphasising the relationships between component traits (phenology and grain yield) and resistance levels. These findings facilitate the development of pigeonpea varieties exhibiting improved resistance to pod borer.

2 Materials and methods

2.1 Plant material

This investigation used 146 accessions from the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) minicore collection (Upadhyaya et al., 2006), along with two checks (resistant check ICPL 332WR and susceptible check ICPL 87). Seed material was procured from the ICRISAT Genebank (https://genebank.icrisat.org/IND/Passport?Crop=Pigeonpea&Location=Passport&mc=Yes).

2.2 Field experiment and phenotyping for pod borer resistance and component traits

Phenotypic screening of 146 accessions, including the susceptible check ICPL 87 and the resistant check ICPL 332WR, was performed using a randomised block design with three replicates during Rainy 2007 (S1), Rainy 2008 (S2), and Rainy 2009 (S3) at ICRISAT-Patancheru, Hyderabad. Each plot consisted of four rows with a row spacing of 30 cm and a plant spacing of 10 cm within each row. Plots were separated by a 1-m alley. Five randomly selected plants from each genotype and replication were tagged for recording observations on pod borer under natural infestation at the maturity stage. PBR was evaluated using a visual damage score on a scale of 1–9, where 1 indicates almost no damage (resistant) and 9 represents severe damage (highly susceptible) (Supplementary Figure 1), during the podding stage (Sujana et al., 2008). This was assessed alongside component traits such as days to 50% flowering (DF), days to maturity (DM), and grain yield (GY). DF and DM were recorded on a per-plant basis. The following pigeonpea descriptors (IBPGR and ICRISAT, 1993) were used to record the GY (g) per plant on five randomly selected representative plants per plot. The four best minicore accessions were subjected to a pod bioassay with artificial third-instar larvae (Ambidi et al., 2021).

2.3 Statistical analyses of phenotypic data

The statistical analysis of the phenotypic data was performed using RStudio version 4.3.1 (http://www.rstudio.com/). The "FactoMineR" package in R was used to perform Pearson's correlation on replicated data from three seasons (Lê et al., 2008). The R package "phenotype" was used to calculate the best linear unbiased predictions (BLUPs) (Piepho et al., 2008).

2.4 DNA extraction and whole-genome resequencing

Genomic DNA was isolated from young leaves using the NucleoSpin® 96 Plant II Kit (Macherey-Nagel), Düren, Nordrhein-Westfalen, Germany. The quality was assessed through 0.8% agarose gel electrophoresis, and the amount was quantified using a Qubit® 2.0 fluorometer (Thermo Fisher Scientific Inc.,Waltham, Massachusetts, USA) (Pandey et al., 2020). Libraries with a 500-bp insert size were generated for all samples for WGRS, as detailed in Varshney et al. (2017). The fragments with insert sizes of approximately 500 bp were removed following separation on an agarose gel and then amplified by PCR. Furthermore, each library was subjected to sequencing on the Illumina HiSeq 2500 to generate paired-end reads. The raw reads were subjected to quality check using FastQC v0.11.8 and Trimmomatic v0.39; poor reads (Phred score < 30, read length < 35 bp) and adaptor exhibiting contamination were eliminated, resulting in high-quality reads. Furthermore, high-quality reads were aligned to the improved reference assembly (Cajca.Asha_v2.0) (Garg et al., 2022) using BWA version 0.5.9 (Li et al., 2009) with the standard parameters. SNP calling was conducted using GATK v.3.7 (McKenna et al., 2010). Biallelic SNPs exhibiting less than 20% missing calls and a minor allelic frequency cut-off of 5% and 50% heterozygosity were utilised for further analysis.

2.5 Linkage disequilibrium decay, GWAS analysis, and candidate gene identification

Linkage disequilibrium (LD) was analysed using TASSEL 5.0. The default parameters of PopLDdecay 3.4.2 were employed to calculate the decay of LD with physical distance. GWAS analysis was performed using 4,99,980 polymorphic SNPs and three seasons of pooled phenotyping data recorded on PBR and component traits. In our study, GWAS analysis was conducted employing four models—MLM, CMLM, fixed and random model circulating probability unification (FarmCPU), and settlement of MLM under progressively exclusive relationship (SUPER)—utilising R/GAPIT 4.3.1. The "Bonferroni correction" p-value threshold (<1.00004E−07) was implemented to remove false associations, and only MTAs with a phenotypic variance explained (PVE) >0% were considered (Supplementary Table 1). However, it was found that significant MTAs were only identified using the FarmCPU and SUPER models, which provided the most reliable and statistically significant results, so we considered the MTAs from these two models for downstream analysis. The physical position of significant MTAs with associated traits was used to mine candidate genes in these regions on the pigeonpea reference genome assembly v2.0 (Garg et al., 2022). We only considered genes where significant MTAs were located (genic and non-genic

10.3389/fpls.2025.1630435

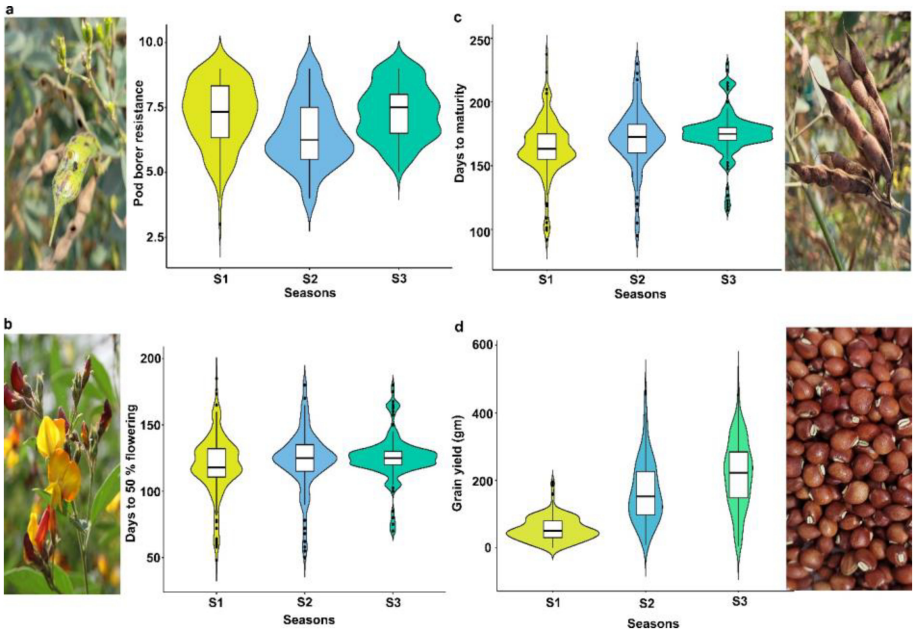


FIGURE 1
Phenotypic variation in minicore accessions for PBR and component traits. Violin plot showing variation for **(a)** pod borer resistance (PBR), **(b)** days to 50% flowering, **(c)** days to maturity, and **(d)** grain yield (g) traits consecutively evaluated for three seasons: S1, Rainy 2007; S2, Rainy 2008; and S3, Rainy 2009.

regions) based on variant annotation and the prediction of SNP effects using the open-source SNPEff-4.3T program.

3 Results

3.1 Phenotypic variation, heritability, and correlation for PBR and component traits

Phenotypic evaluation for PBR and component traits showed significant variation among the minicore accessions. A symmetric distribution was observed for most of the traits (Figure 1). PBR score showed differences across seasons (3–9 in S1, 4–9 in S2, and 5–9 in S3). The average scores recorded were 7.3, 6.5, and 7.3 in S1, S2, and S3, respectively. Similarly, accessions revealed a wide range of variation for component traits: DF (48–185 in S1, 50–180 in S2, and 70–180 in S3), DM (90–245 in S1, 95–230 in S2, and 115–230 in S3), and GY (3–287 g plant^{-1} in S1, 8–463 g plant^{-1} in S2, and 6–481 g plant^{-1} in S3). Across seasons, all traits showed larger phenotypic coefficient of variation (PCV) and genotypic coefficient of variation (GCV) (>10%). Broad-sense heritability (h^2) averaged 54% for PBR, 92% for DF, 96% for DM, and 55% for GY (Table 1). Following the replicated multi-season field evaluation results, 19 best lines were selected and screened using a lab-based pod bioassay. The gain % of the resistant check (ICPL332WR; score 6) compared to the susceptible check (ICPL 87; score 9) was assessed. Among the best lines, four—CP 10503, ICP 655, ICP 9691, and ICP 9655 (scoring between 4 and 5)—showed the least damage rating and low larval weight. Pearson's correlation test was performed to determine the phenotypic correlation between PBR and component traits. A total of six possible correlations were observed, with three pairs (one positive and two negative). Correlation discussion revealed a strong positive correlation between DM and DF (r = 0.97), whereas GY and PBR had the highest negative correlation (r = −0.55), which was significant at the 0.001 level. Other correlations were not statistically significant ($p > 0.05$) (Figure 2).

3.2 LD decay and genome-wide association study analysis

A total of 4,99,980 filtered SNP markers with a genotype call rate >0.80 and a minor allele frequency (MAF) of >5% were utilised for downstream analysis. The overall LD decay across the 11 chromosomes was 39.5 kbp on average (Figure 3). For the GWAS analysis, high-quality phenotyping data (three seasons pooled phenotyping data) for PBR and component traits were used along with 4,99,980 polymorphic SNPs. Two models (SUPER and FarmCPU) identified 14 significant MTAs (eight PBR, one DF, two DM, and three GY) for four traits, explaining 0.05%–28.1% phenotypic variation with a *p*-value range of 1.10E−13 to 9.66E−09 for pooled data (Table 2). Eight significant MTAs were identified for PBR on chromosomes CcLG02, CcLG04, CcLG05, CcLG07, and CcLG11, with PVE ranging from 0.05% to 5.57% (Figure 4). For DF, one MTA was detected (CcLG11_38698041) on the same chromosome (CcLG11) with PVE of 28.1%. On chromosome CcLG04, two MTAs (CcLG04_38227177 and CcLG04_7181399) for DM were detected, explaining 1.07%–2.74% PVE. Three MTAs were identified for GY on chromosomes CcLG02, CcLG05, and CcLG07, accounting for 0.49%–1.72% phenotypic variance (Figure 5). Based on pooled phenotyping data, a representative set

TABLE 1 Mean, range, and variability components in minicore accessions for PBR and component traits across three seasons.

Traits	Seasons	Range	Grand mean	CD (5%)	SE (±)	GCV	PCV	h^2 (%)
DF	S1	48–185	118.5	14.8	5.3	17.9	19.5	80
	S2	50–180	123.7	7.5	2.7	18.3	18.5	97
	S3	70–180	125.6	3	1.1	13.6	13.6	99
DM	S1	90–245	162.7	6.1	2.2	13.7	13.9	97
	S2	95–230	171.3	10.6	3.8	14	14.3	95
	S3	115–230	175.5	4	1.4	10.1	10.2	98
PBR	S1	3–9	7.3	1.8	0.7	13.3	20.6	42
	S2	4–9	6.5	1.6	0.6	16.7	21	63
	S3	5–9	7.3	1.5	0.5	12.3	16	58
GY (g)	S1	3–287	124.1	144	34.9	57.4	76.9	56
	S2	8–463	168.5	154.9	55.4	43.7	63.8	47
	S3	6–481	216.7	129	46.2	40.4	50.4	64

S1, Rainy 2007; S2, Rainy 2008; S3, Rainy 2009; DF, days to 50% flowering; DM, days to maturity; PBR, pod borer resistance; GY, grain yield; SE, standard error; CD, critical difference; GCV, genotypic coefficient of variation; PCV, phenotypic coefficient of variation; h^2, broad-sense heritability.

of minicore accessions, exhibiting variability for PBR and component traits, were selected to *in silico* validate the SNPs associated with the significant MTAs. Among the 14 detected MTAs, five showed polymorphism, including two for PBR (CcLG04_35844765 and CcLG07_10581882) (Supplementary Figure 2) and three for component traits (one for DF, CcLG01_38698041; one for DM, CcLG04_7181399; and one for GY, CcLG04_25609089) in the minicore accessions (Supplementary Figures 3–5). These results indicate that the identified SNPs could be used to develop allele-specific markers for MAS, helping develop pigeonpea cultivars with improved resistance to *H. armigera*.

3.3 Putative genes associated with MTAs

The putative genes linked with the 14 significant MTAs identified for PBR and component traits were examined by analysing their location (genic and non-genic), effects, and functions (Table 2). Among these, 12 were detected in the intergenic regions, while one each was found in the exonic, 5′ UTR, and synonymous variant regions. Notably, MTA CcLG05_29876072, located in the 5′ UTR of gene (*Cc_11847*) on CcLG05 chromosome, was associated with PBR. Additionally, MTA CcLG11_49001007, located in the synonymous variant region of gene (*Cc_23491*) on CcLG11 chromosome, was also associated with PBR. Furthermore, the remaining significant MTAs associated with PBR, DF, DM, and GY were found in intergenic regions.

4 Discussion

Pigeonpea is a major grain legume that draws worldwide interest for its important contribution to nutritional and food security. However, *H. armigera* results in substantial losses in yield, posing a severe obstacle in pigeonpea cultivation (Sharma et al., 2022). Although significant breeding efforts have been made, the development of resistant varieties remains difficult due to the complex inheritance of resistance traits and the lack of genetic variation in cultivated germplasm (Volp et al., 2023; Karrem et al., 2025). Besides, transgenic approaches have potential legal issues and public acceptability concerns in India (Rakesh and Ghosh, 2024). Moreover, conventional pest management practices often lack sustainable solutions due to resistance to pesticides and environmental issues. Therefore, investigation of pigeonpea germplasm in the primary gene pool and crop wild relatives is an effective option since several *Cajanus* species exhibit higher resistance to *H. armigera* (Sharma et al., 2022; Singh et al., 2022). However, most cultivated genotypes showed low to moderate levels of resistance to *H. armigera*, evidenced by the screening of nearly 14,000 pigeonpea accessions (Reed and Lateef, 1990). Several investigations have reported that a few accessions of the wild progenitor of pigeonpea have exhibited high levels of resistance to *H. armigera* (Green et al., 2006; Sharma et al., 2009, 2022). It is important to understand that the trait is substantially influenced by genetic and environmental factors; therefore, relying solely on phenotypic screening for selection is insufficient. Furthermore, understanding the genetic basis of resistance to *H. armigera* can provide opportunities for developing resistant varieties. Our investigation on minicore accessions reported a broad range of variation in PBR and component traits. The h^2 values of PBR, DF, DM, and GY were 54%, 92%, 96%, and 55%, respectively, indicating that a significant portion of the variation is attributable to distinct genotypes. The correlation analysis revealed a strong positive correlation between DM and DF, indicating that days to flowering could act as an index for maturity classification in pigeonpea. Previous investigations have shown similar findings for a correlation between DM and DF (Singh et al., 1995). Compared to other traits, GY and PBR had the strongest negative association (r = −0.55), which shows

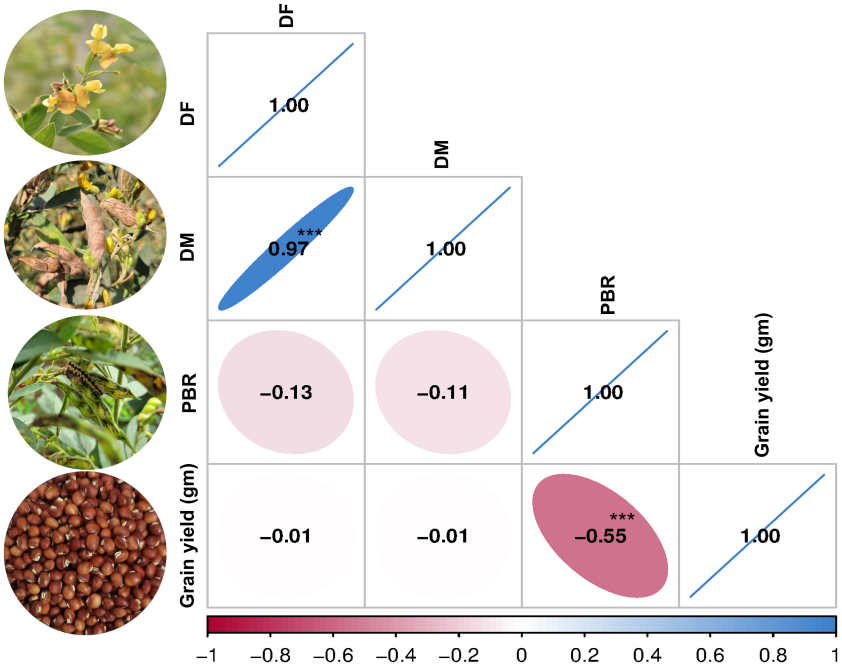


FIGURE 2
Pearson's correlation matrix. Correlation between pod borer resistance (PBR) and component traits (***significant at 0.001 level).

the highly influential nature of the trait. This indicates that a higher PBR score tends to be associated with lower GY, or vice versa. Furthermore, environmental factors such as excessive rainfall during sowing in S1, delayed planting, and variations in day and night temperatures throughout the reproductive stages likely contributed to the lower yield in S1, despite similar pod borer scores in other seasons. Previous studies have also reported that delayed sowing reduces yield in pigeonpea (Arunkumar et al., 2018).

MAS is an promising approach for accelerating the development of insect pest-resistant varieties. It facilitates the development of multi-trait resistant varieties by pyramiding different resistance genes to target insects, which is not possible with traditional breeding due to similar expression of phenotype (Sharma and Crouch, 2004). Utilising WGRS data along with precise phenotypic variability could help identify accessions with rare variants that may be potentially linked with important traits, such as resistance to *H. armigera*. In GWAS, determining the pattern of LD is important since it influences the resolution and magnitude of the association analysis. Our analysis showed an average LD decay at 39.5 kbp. The rapid LD decay indicates a minimal extent of long-range LD among the minicore accessions. A previous study reported genome-wide LD decay at 118 kb (Megha et al., 2024). GWAS minimises the two primary constraints of traditional linkage mapping, such as limited allelic diversity and insufficient genetic resolution (Huang and Han, 2014). Due to its high resolution and low cost in sequencing/genotyping, GWAS analysis has successfully dissected important traits in pigeonpea, including flowering-related traits (Kumar et al., 2022) and antioxidant activity (Megha et al., 2024). The main concern for GWAS is to minimise false positives, mostly due to population structure and familial relatedness (Kaler et al., 2020). Although single-locus models overcome this issue by including the two confounding factors as covariates, over-fitting in a model usually leads to false negatives, which could eliminate valuable loci (Price et al., 2006). In this context, multi-locus models provide an alternative way for reducing false negatives (Zhang et al., 2019). Multi-locus GWAS models, such as the SUPER and FarmCPU methods, improve statistical power but minimise false positives. The SUPER model offers greater computational power and requires less computing than earlier models. However, it extracts a small number of SNPs termed

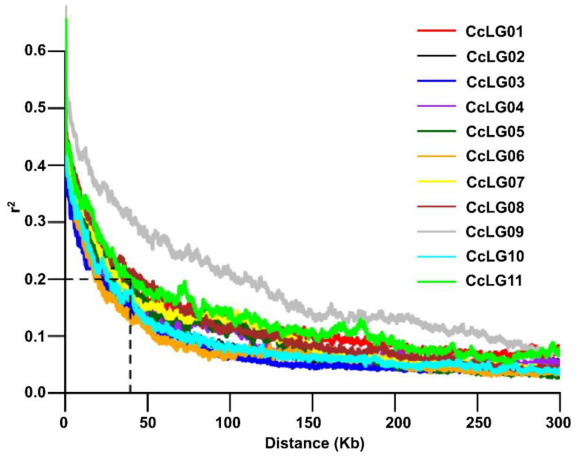


FIGURE 3
Estimated linkage disequilibrium decay (LD decay). LD decay for each chromosome, with r^2 ~ 0.2 at 39.5 kb.

TABLE 2 Significant MTAs detected for PBR and component traits using multi-locus models with predictive gene variants and their functions.

Traits	Significant MTAs	Models	Chr	Position (bp)	*p*-Value	PVE (%)	Gene variants	Alleles	Gene ID	Functional annotation
PBR	CcLG02_16059144	SUPER	CcLG02	16,059,144	8.56E−09	1.52	DGV	G/A	*Cc_3152*	*Probable carboxylesterase 15*
PBR	CcLG04_31940898	SUPER	CcLG04	31,940,898	6.40E−09	1.30	DGV	T/C	*Cc_10409*	*Microtubule-associated protein 5*
PBR	CcLG04_35844765	SUPER	CcLG04	35,844,765	9.55E−09	5.57	UGV	G/A	*Cc_10514*	*Phospholipase D delta*
PBR	CcLG05_29876072	SUPER	CcLG05	29,876,072	2.21E−08	3.35	5′ UTR	A/T	*Cc_11847*	*Uncharacterized protein LOC100796483*
PBR	CcLG05_10183191	SUPER	CcLG05	10,183,191	9.66E−09	0.31	UGV	C/A	*Cc_11087*	*L-type lectin-domain containing receptor kinase IX.1*
PBR	CcLG05_9948484	SUPER	CcLG05	9,948,484	2.13E−08	0.15	UGV	C/A	*Cc_11074*	*Hypothetical protein GLYMA_14G106500*
PBR	CcLG07_10581882	SUPER	CcLG07	10,581,882	1.95E−11	0.05	UGV	G/A	*Cc_15562*	*Omega-hydroxypalmitate O-feruloyl transferase*
PBR	CcLG11_49001007	SUPER	CcLG11	49,001,007	7.86E−08	4.48	SV	T/C	*Cc_23491*	*FAR1-RELATED SEQUENCE 4*
DF	CcLG11_38698041	SUPER	CcLG011	38,698,041	5.96E−08	28.11	UGV	C/T	*Cc_23683*	*Hypothetical protein LR48_Vigan05g175700*
DM	CcLG04_38227177	FarmCPU	CcLG04	38,227,177	5.53E−08	1.07	UGV	G/T	*Cc_10599*	*RING-H2 finger protein ATL7*
DM	CcLG04_7181399	FarmCPU	CcLG04	7,181,399	1.74E−11	2.74	UGV	G/A	*Cc_9357*	*Leucine-rich repeat receptor-like protein kinase*
GY	CcLG02_25609089	FarmCPU	CcLG02	25,609,089	1.10E−13	0.49	UGV	C/T	*Cc_3417*	*Glutaredoxin-C5, chloroplastic*
GY	CcLG05_16338260	FarmCPU	CcLG05	16,338,260	1.75E−12	1.72	DGV	C/G	*Cc_11279*	*Secretory carrier-associated membrane protein*
GY	CcLG07_9297796	FarmCPU	CcLG07	9,297,796	1.22E−11	0.97	UGV	C/T	*Cc_15498*	*Hypothetical protein GLYMA_16G039300*

MTAs, marker–trait associations; Chr, chromosome; PBR, pod borer complex resistance; DF, days to flowering; DM, days to maturity; GY, grain yield; UGV, upstream gene variant; DGV, downstream gene variant; 5′ UTR, 5 prime UTR variant; SV, synonymous variant; MGV, missense gene variant; PVE, phenotypic variance explained; SUPER, settlement of MLM under progressively exclusive relationship; FarmCPU, fixed and random model circulating probability unification.

pseudo-quantitative trait nucleotide (QTN) to determine kinship (Wang et al., 2014). Moreover, "FarmCPU" is a novel multi-locus model that is computationally powerful and efficiently controls false negatives and false positives. Two multi-locus methods (SUPER and FarmCPU) were included in the current investigation to identify significant MTAs for PBR and component traits. GWAS analysis identified 14 significant MTAs linked to four traits, including eight for PBR, three for GY, two for DM, and one for DF. For the DF trait, one MTA was detected on chromosome CcLG11, accounting for the highest phenotypic variation of 28.1%. Most of the identified MTAs exhibited smaller phenotypic variation % and lower *p*-values. This finding suggests that these traits are controlled by multiple genes with minor effects, reflecting complex genetic architecture, and are also influenced by environmental factors. The statistical power of association mapping could be substantially improved by increasing the population size (Liu et al., 2021). The MTAs detected for PBR and component traits in our study were not reported previously and seem to indicate novel genetic loci in pigeonpea. Thus, the SNPs associated with MTAs offer the possibility of additional validation in diverse collections and may be utilised for early generation selection in breeding programs.

A total of 14 significant MTAs for four traits were detected and linked with putative genes. One MTA for PBR was found on chromosome 2 (CcLG02_16059144) linked to the *Cc_3152* gene encoding a *probable carboxylesterase 15* enzyme that catalyses the conversion of carboxylic esters and water into alcohol and carboxylate. In plants, it is involved in defence, development, and secondary metabolism (Palayam et al., 2024). In tobacco, this gene (NbCXE) is involved in host defence responses against Tobacco mosaic virus (TMV) infection (Guo and Wong, 2020). Similarly, another MTA (CcLG04_31940898) was identified for PBR encoding *microtubule-associated protein 5*, which plays a key role in cell division, cell proliferation, and cell morphology. In *Arabidopsis*, the microtubule-binding protein (*TGNap1*) facilitates the secretion of antimicrobial proteins, important for defence against phytopathogens (Bhandari et al., 2023). The MTA detected for PBR (CcLG11_49001007) in the exonic region of gene *Cc_23491*, which encodes *FAR1-RELATED SEQUENCE*, is a

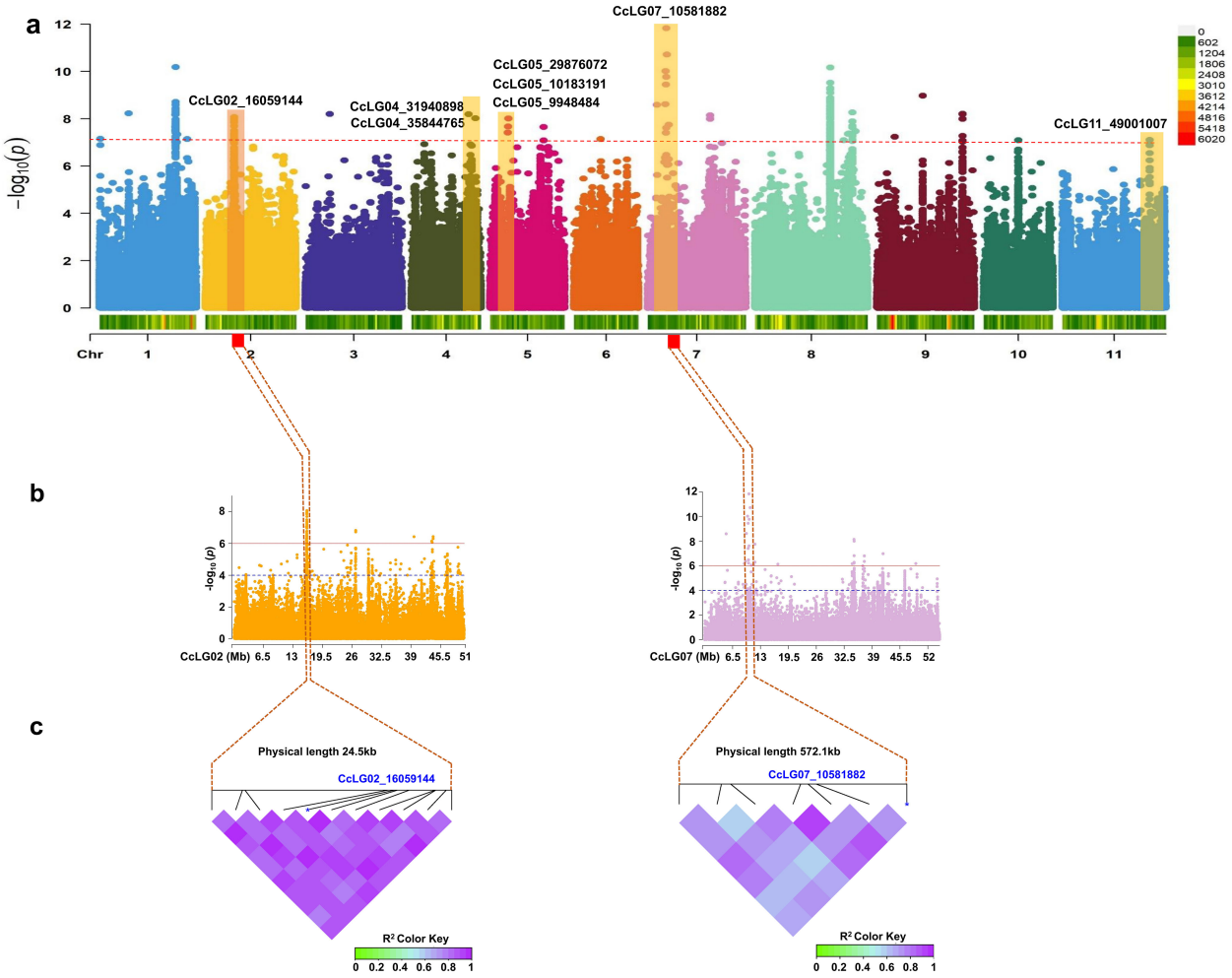


FIGURE 4
Genome-wide association study (GWAS) for pod borer resistance (PBR) trait. **(a)** Manhattan plot illustrating significant marker–trait association (MTA) for PBR trait. Only highly statistically significant MTAs at peak were considered. **(b)** Association for the significant SNPs on chromosomes CcLG02 and CcLG07. The interval of association was determined to lie between 24.5 kb downstream of the significant MTA (CcLG02_16059144) on CcLG02 and 572.1 kb downstream of the significant MTA (CcLG07_10581882) on CcLG07. **(c)** Linkage disequilibrium heatmaps for the association region for PBR on chromosomes CcLG02 and CcLG07. Bonferroni correction threshold of *p*-value (<1.00004E−07) was implemented to detect significant associations.

light signalling factor pair with *FAR-RED ELONGATED HYPOCOTYL 3* to regulate plant immunity by integrating chlorophyll biosynthesis with the salicylic acid (SA) signalling pathway in *Arabidopsis* (Wang et al., 2015). For PBR, three more MTAs were detected and associated with *Cc_10514*, *Cc_11087*, and *Cc_15562*. Gene *Cc_10514* encodes *phospholipase D delta*, a protein that is involved in basal defence and non-host resistance to powdery mildew fungi in *Arabidopsis* (Pinosa et al., 2013). *Cc_11087* encodes an L-type lectin domain-containing receptor kinase IX, involved in self/non-self-surveillance and plant resistance. The homologues of these receptors in *Nicotiana benthamiana* and *Solanum lycopersicum* have the same role in defence against *Phytophthora* (Wang et al., 2015), and an MTA identified for PBR on chromosome 7 (CcLG07_10581882) associated with the *Cc_15562* gene encoding *omega-hydroxypalmitate O-feruloyl transferase* has a role in suberin biosynthesis. Suberin is synthesised in plant wound tissues to prevent pathogen infection (Molina et al., 2009). For PBR, two MTAs (CcLG05_29876072 and CcLG05_9948484) were detected in exonic (*Cc_11087* gene) and intergenic (*Cc_11074* gene) regions. These were predicted to encode an *uncharacterized protein* and a *hypothetical protein GLYMA_1*.

The MTA (CcLG04_38227177) identified for DM lies in the *Cc_10599* gene, which encodes a *RING-H2 finger protein*, which is important for seed development in *Arabidopsis* (Xu and Quinn Li, 2003). The MTA (CcLG04_7181399), identified for DM, is associated with the *Ca_00148* gene, which encodes a *leucine-rich repeat receptor-like protein kinase*. This protein is an important membrane-bound regulator of abscisic acid (ABA) early signalling in *Arabidopsis*, and ABA is involved in seed maturation (Osakabe et al., 2005). For GY, the *CcLG05_25609089* MTA was associated

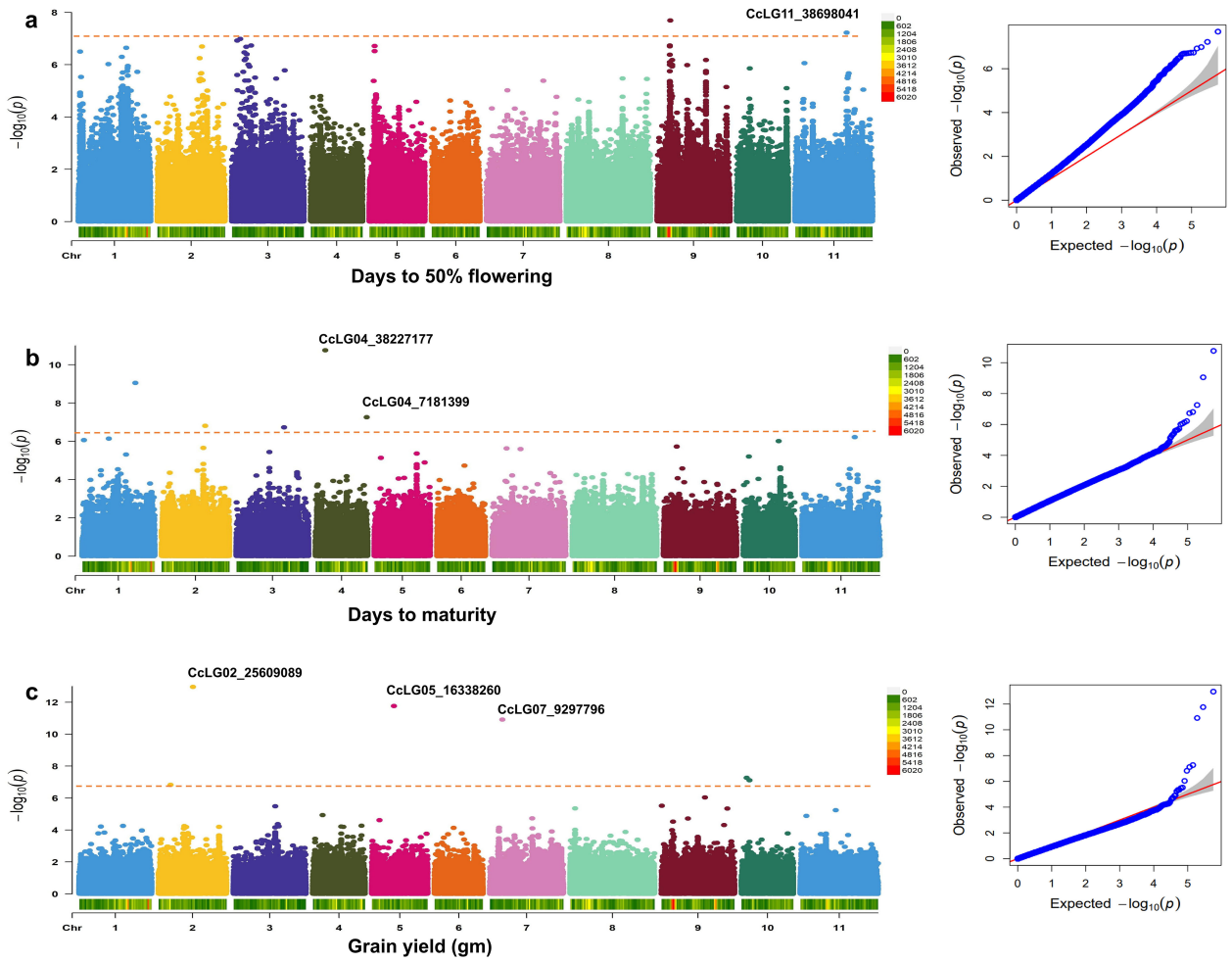


FIGURE 5
Genome-wide association study (GWAS) for component traits. Manhattan and quantile–quantile (QQ) plots for **(a)** days to 50% flowering, **(b)** days to maturity, and **(c)** grain yield. Bonferroni correction threshold of *p*-value (<1.00004E−07) was implemented to detect significant associations.

with the *Cc_3417* gene. *Cc_3417* encodes a *glutaredoxin-C5 chloroplastic protein*. Overexpression of a *CPYC-type glutaredoxin* was shown to increase grain weight in rice (Liu et al., 2019). Another MTA (CcLG05_16338260) is present in the intergenic region of the *Cc_11279* gene, encoding for the *secretory carrier-associated membrane protein*. Karnik et al. (2013) demonstrated that secretory carrier membrane proteins (SCAMPs) are involved in the secretion of defence proteins, including protease inhibitors and toxins, in *Arabidopsis thaliana*. These proteins have been shown to inhibit insect feeding or growth. However, further validation of the identified MTAs is required across varying genetic backgrounds. This provides deeper insights into the genetic control of resistance mechanisms, along with the potential to develop effective markers (Thakur et al., 2025). Additionally, gene editing innovations offer promising tools for validating and modifying the candidate genes identified by GWAS. It enables the precise knock-in or knockout of specific genes, providing clear evidence of their role in resistance. The integration of detected genes and SNPs associated with MTAs through molecular breeding or genetic modification could provide an effective approach for developing *H. armigera*-resistant cultivars.

5 Conclusion

Pod borer, *H. armigera*, is one of the most damaging pests in pigeonpea production. Various methods have been employed for controlling this pest, but have exhibited limited success. Phenotypic data on PBR and component traits, along with genotypic data from the WGRS, were used to identify 14 significant MTAs. These significant MTAs had 0.05%–28.1% phenotypic variation with a *p*-value range of 1.10E−13 to 9.66E−09. MTA for DF (CcLG11_38698041) on chromosome (CcLG11) had the highest PVE of 28.1%. Furthermore, we identified that important genes that encode *probable carboxylesterase 15* (*Cc_3152*), *microtubule-associated protein 5* (*Cc_10409*), and *FAR1-RELATED SEQUENCE* (*Cc_23491*) have been associated with plant defence responses and the regulation of plant immunity. These putative genes can be helpful for the identification of molecular targets, providing insight into the biological pathways that underlie the traits of interest and facilitating understanding of the genetic basis of complex traits. The importance of these genomic regions for future studies will help to understand the *H. armigera*-resistant mechanism, along with finding functional markers. Notably,

further lab-based pod bioassay screening identified four minicore accessions—ICP 10503, ICP 655, ICP 9691, and ICP 9655—which showed moderate resistance. The resistant genotypes, significant MTAs, and putative genes identified in this investigation have the potential to be utilised in the development of pod borer-resistant pigeonpea cultivars.

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

AM: Validation, Data curation, Writing – original draft, Methodology, Formal Analysis, Visualization. RM: Methodology, Writing – review & editing, Data curation. VS: Software, Visualization, Validation, Writing – review & editing, Formal Analysis, Methodology, Data curation. SM: Methodology, Data curation, Writing – review & editing, Formal Analysis. JJ: Investigation, Formal Analysis, Writing – review & editing, Validation, Data curation, Methodology. ShG: Data curation, Writing – review & editing, Formal Analysis. SuG: Software, Formal Analysis, Writing – review & editing. PG: Formal Analysis, Writing – review & editing, Data curation. ND: Visualization, Writing – review & editing, Software. SR: Writing – review & editing. RR: Writing – review & editing. PG: Writing – review & editing. HS: Writing – review & editing, Investigation, Conceptualization, Supervision, Resources, Project administration, Data curation, Methodology. MKP: Validation, Resources, Project administration, Investigation, Writing – review & editing, Methodology, Funding acquisition, Supervision, Conceptualization.

Funding

The author(s) declare financial support was received for the research and/or publication of this article. The authors gratefully acknowledge the financial assistance provided by the ICAR-ICRISAT collaborative project, India; Department of Biotechnology, Government of India; Global Initiative project VACS (Vision for Adapted Crops and Soils); and Tropical Legumes Project TLII from the Bill and Melinda Gates Foundation, which provided funding support in parts to the present study.

Acknowledgments

AM and VS acknowledges Chaudhary Charan Singh University (CCSU), Meerut, for collaborating with ICRISAT. The authors are thankful to ICRISAT Genebank for providing seed material.

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.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2025.1630435/full#supplementary-material

References

Ambidi, V., Bantewad, S., Prasad Mishra, S., Hingane, A., and Jaba, J. (2021). Morpho-biochemical parameters associated with resistance to pod borer complex of pigeonpea. *Pak. J. Zool.* 54, 405–411. doi: 10.3390/ijms22158327

Arunkumar,, Dhanoji, M. M., and Meena, M. K. (2018). Phenology and productive performance of pigeon pea as influenced by date of sowing. *J. Pharmacogn. Phytochem.* 7, 266–268.

Bhandari, D. D., Ko, D. K., Kim, S. J., Nomura, K., He, S. Y., and Brandizzi, F. (2023). Defense against phytopathogens relies on efficient antimicrobial protein secretion mediated by the microtubule-binding protein TGNap1. *Nat. Commun.* 14, 6357. doi: 10.1038/s41467-023-41807-4

FAO (2024). *FAO Statistic Division*. Available online at: www.fao.org (Accessed November 14, 2024).

Garg, V., Dudchenko, O., Wang, J., Khan, A. W., Gupta, S., Kaur, P., et al. (2022). Chromosome-length genome assemblies of six legume species provide insights into genome organization, evolution, and agronomic traits for crop improvement. *J. Adv. Res.* 42, 315–329. doi: 10.1016/j.jare.2021.10.009

Ghosh, G., Ganguly, S., Purohit, A., Chaudhuri, R. K., Das, S., and Chakraborti, D. (2017). Transgenic pigeonpea events expressing *Cry1Ac* and *Cry2Aa* exhibit resistance to *Helicoverpa armigera*. *Plant Cell Rep.* 36, 1037–1051. doi: 10.1007/s00299-017-2133-0

Green, P. W. C., Sharma, H. C., Stevenson, P. C., and Simmonds, M. S. J. (2006). Susceptibility of pigeonpea and some of its wild relatives to predation by *Helicoverpa armigera*: implications for breeding resistant cultivars. *Aust. J. Agric. Res.* 57, 831–836. doi: 10.1071/AR05281

Gudi, S., Halladakeri, P., Singh, G., Kumar, P., Singh, S., Alwutayd, K. M., et al. (2024). Deciphering the genetic landscape of seedling drought stress tolerance in wheat (*Triticum aestivum* L.) through genome-wide association studies. *Front. Plant Sci.* 15, 1351075. doi: 10.3389/fpls.2024.1351075

Guo, S., and Wong, S. M. (2020). A conserved carboxylesterase inhibits tobacco mosaic virus (TMV) accumulation in *Nicotiana benthamiana* plants. *Viruses* 12, 195. doi: 10.3390/v12020195

Halladakeri, P., Gudi, S., Akhtar, S., Singh, G., Saini, D. K., Hilli, H. J., et al. (2023). Meta-analysis of the quantitative trait loci associated with agronomic traits, fertility restoration, disease resistance, and seed quality traits in pigeonpea (*Cajanus cajan* L.). *Plant Genome* 16, e20342. doi: 10.1002/tpg2.20342

Huang, X., and Han, B. (2014). Natural variations and genome-wide association studies in crop plants. *Annu. Rev. Plant Biol.* 65, 531–551. doi: 10.1146/annurev-arplant-050213-035715

IBPGR and ICRISAT (1993). *Descriptors for pigeonpea [*Cajanus cajan *(L.) Millsp.]* (International Board of Plant Genetic Resources: Rome, Italy; International Crops Research Institute for Semi-Arid Tropics: Patancheru, India,), 31p.

Jaba, J., Vashisth, S., Golla, S. K., and Mishra, S. P. (2023). Effect of different Sowing Windows on Major Insect Pests and Host Plant Resistance to Pod Borer, *Helicoverpa armigera* in Pigeonpea (*Cajanus cajan* (L.) Millsp.). *Pak. J. Zool.*, 56, 1–10. doi: 10.17582/journal.pjz/20210320070322

Kaler, A. S., Gillman, J. D., Beissinger, T., and Purcell, L. C. (2020). Comparing different statistical models and multiple testing corrections for association mapping in soybean and maize. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.01794

Kambrekar, D. N. (2016). Management of legume pod borer, *Helicoverpa armigera* with host plant resistance. *Legume Genom. Genet.* 29, 157–171. doi: 10.5376/lgg.2016.07.0005

Kang, Y., Torres-Jerez, I., An, Z., Greve, V., Huhman, D., Krom, N., et al. (2019). Genome-wide association analysis of salinity responsive traits in *Medicago truncatula*. *Plant Cell Environ.* 4, 1513–1531. doi: 10.1111/pce.13508

Karnik, R., Grefen, C., Bayne, R., Honsbein, A., Köhler, T., Kioumourtzoglou, D., et al. (2013). Arabidopsis Sec1/Munc18 protein SEC11 is a competitive and dynamic modulator of SNARE binding and SYP121-dependent vesicle traffic. *Plant Cell* 25, 1368–1382. doi: 10.1105/tpc.112.108506

Karrem, A., Haveri, R. V., Yogendra, K., Prabhuraj, A., HanChinal, S., Kalyan, A., et al. (2025). Understanding resistance mechanisms in crop wild relatives (CWRs) of pigeonpea (*Cajanus cajan* L.) against pod borer *Helicoverpa armigera* (Hub.). *Genet. Resour. Crop Evol.* 72, 7577–7597. doi: 10.1007/s10722-025-02392-1

Kinhoégbè, G., Djèdatin, G., Saxena, R. K., Chitikineni, A., Bajaj, P., Molla, J., et al. (2022). Genetic diversity and population structure of pigeonpea (*Cajanus cajan* [L.] Millspaugh) landraces grown in Benin revealed by Genotyping-By-Sequencing. *PloS One* 17, 271565. doi: 10.1371/journal.pone.0271565

Kumar, K., Anjoy, P., Sahu, S., Durgesh, K., Das, A., Tribhuvan, K. U., et al. (2022). Single trait versus principal component based association analysis for flowering related traits in pigeonpea. *Sci. Rep.* 12, 10453. doi: 10.1038/s41598-022-14568-1

Lê, S., Josse, J., and Husson, F. (2008). FactoMineR: An R package for multivariate analysis. *J. Stat. Software* 25, 1–18. doi: 10.18637/jss.v025.i01

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

Liu, C., Ding, X., Wu, Y., Zhang, J., Huang, R., Li, X., et al. (2024). Chromosome-scale reference genome of an ancient landrace: unveiling the genetic basis of seed weight in the food legume crop pigeonpea (*Cajanus cajan*). *Hortic. Res.* 11, 201. doi: 10.1093/hr/uhae201

Liu, S., Fu, H., Jiang, J., Chen, Z., Gao, J., Shu, H., et al. (2019). Overexpression of a CPYC-type glutaredoxin, *OsGrxC2.* 2, causes abnormal embryos and an increased grain weight in rice. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.00848

Liu, Y., Hu, G., Zhang, A., Loladze, A., Hu, Y., Wang, H., et al. (2021). Genome-wide association study and genomic prediction of *Fusarium* ear rot resistance in tropical maize germplasm. *Crop J.* 9, 325–341. doi: 10.1016/j.cj.2020.08.008

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

Megha,, Singh, N., Sharma, M., Langyan, S., and Kumar Singh, N. (2024). Genome wide association study of antioxidant activity in pigeonpea germplasm. *Discov. Food.* 4, 82. doi: 10.1007/s44187-024-00160-1

Molina, I., Li-Beisson, Y., Beisson, F., Ohlrogge, J. B., and Pollard, M. (2009). Identification of an Arabidopsis feruloyl-coenzyme A transferase required for suberin synthesis. *Plant Physiol.* 151, 1317–1328. doi: 10.1104/pp.109.144907

Osakabe, Y., Maruyama, K., Seki, M., Satou, M., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2005). Leucine-rich repeat receptor-like kinase1 is a key membrane-bound regulator of abscisic acid early signaling in Arabidopsis. *Plant Cell* 17, 1105–1119. doi: 10.1105/tpc.104.027474

Palayam, M., Yan, L., Nagalakshmi, U., Gilio, A. K., Cornu, D., Boyer, F. D., et al. (2024). Structural insights into strigolactone catabolism by carboxylesterases reveal a conserved conformational regulation. *Nat. Commun.* 15, 6500. doi: 10.1038/s41467-024-50928-3

Pandey, M. K., Gangurde, S. S., Sharma, V., Pattanashetti, S. K., Naidu, G. K., Faye, I., et al. (2020). Improved genetic map identified major QTLs for drought tolerance-and iron deficiency tolerance-related traits in groundnut. *Genes* 12, 37. doi: 10.3390/genes12010037

Piepho, H. P., Möhring, J., Melchinger, A. E., and Büchse, A. (2008). BLUP for phenotypic selection in plant breeding and variety testing. *Euphytica* 161, 209–228. doi: 10.1007/s10681-007-9449-8

Pinosa, F., Buhot, N., Kwaaitaal, M., Fahlberg, P., Thordal-Christensen, H., Ellerström, M., et al. (2013). Arabidopsis phospholipase Dδ is involved in basal defense and nonhost resistance to powdery mildew fungi. *Plant Physiol.* 163, 896–906. doi: 10.1104/pp.113.223503

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* 38, 904–909. doi: 10.1038/ng1847

Rakesh, V., and Ghosh, A. (2024). Advancements in genetically modified insect pest-resistant crops in India. *Planta* 260, 86. doi: 10.1007/s00425-024-04511-1

Reed, W., and Lateef, S. S. (1990). *Pigeonpea: pest management in the pigeonpea* (UK: CAB Internat), 349–374, ISBN: .

Saxena, R. K., Kale, S. M., Kumar, V., Parupali, S., Joshi, S., Singh, V., et al. (2017). Genotyping-by-sequencing of three mapping populations for identification of candidate genomic regions for resistance to sterility mosaic disease in pigeonpea. *Sci. Rep.* 7, 1813. doi: 10.1038/s41598-017-01535-4

Seetharamu, P., Swathi, K., Dhurua, S., Suresh, M., Govindarao, S., Sreesandhya, N., et al. (2020). Bioefficacy of chemical insecticides against major sucking insect pests on grain legumes in India-A review. *Legume Res.* 43, 1–7. doi: 10.18805/LR-4074

Shanower, T. G., Romeis, J., and Minja, E. M. (1999). Insect pests of pigeonpea and their management. *Annu. Rev. Entomol.* 44, 77–96. doi: 10.1146/annurev.ento.44.1.77

Sharma, H. C. (2016). Climate change vis-a-vis pest management. Proceedings in conference on national priorities in plant health management. *Tirupati*, 17–25.

Sharma, H. C., and Crouch, J. H. (2004). "Molecular marker-assisted selection: a novel approach for host plant resistance to insects in grain legumes," in *Pulses in new perspective*. Eds. M. Ali, B. B. Singh, S. Kumar and V. Dhar (Indian Society of Pulses Research and Development, Kanpur), 147–173.

Sharma, S., Jaba, J., Rao, P. J., Prasad, S., Gopal, N. T. V. V., Sharma, H. C., et al. (2022). Reaping the potential of wild *Cajanus* species through pre-breeding for improving resistance to pod borer, *Helicoverpa armigera*, in cultivated pigeonpea (*Cajanus cajan* (l.) millsp.). *Biology* 11, 485. doi: 10.3390/biology11040485

Sharma, V., Mahadevaiah, S. S., Latha, P., Gowda, S. A., Manohar, S. S., Jadhav, K., et al. (2024). Dissecting genomic regions and underlying candidate genes in groundnut MAGIC population for drought tolerance. *BMC Plant Biol.* 24, 1–21. doi: 10.1186/s12870-024-05749-3

Sharma, H. C., Sujana, G., and Manohar Rao, D. (2009). Morphological and chemical components of resistance to pod borer, *Helicoverpa armigera* in wild relatives of pigeonpea. *Arthropod Plant Interact.* 3, 151–161. doi: 10.1007/s11829-009-9068-5

Sharma, H. C., Upadhyaya, H. D., Sharma, S. P., Munghate, R. S., Reddy, K. N., and Wadaskar, R. M. (2025). Evaluation of pigeonpea mini-core germplasm collection for resistance to pod borer, *Helicoverpa armigera* (Unpublished).

Singh, N. B., Ariyanayagam, R. P., Gupta, S. C., and Rao, A. N. (1995). Relationship of plant height, days to flowering and maturity to grain yield in short-duration determinate pigeonpea. *Indian J. Genet. Pl. Br.* 55, 1–5.

Singh, G., Gudi, S., Amandeep,, Upadhyay, P., Shekhawat, P. K., Nayak, G., et al. (2022). Unlocking the hidden variation from wild repository for accelerating genetic gain in legumes. *Front. Plant Sci.* 13, 1035878. doi: 10.3389/fpls.2022.1035878

Singh, G., Singh, I., Taggar, G. K., Rani, U., Sharma, P., Gupta, M., et al. (2020). Introgression of productivity enhancing traits, resistance to pod borer and Phytopthora stem blight from *Cajanus scarabaeoides* to cultivated pigeonpea. *PMBP* 26, 1399–1410. doi: 10.1007/s12298-020-00827-w

Sujana, G., Sharma, H. C., and Rao, D. M. (2008). Antixenosis and antibiosis components of resistance to pod borer *Helicoverpa armigera* in wild relatives of pigeonpea. *Int. J. Trop. Insect Sci.* 28, 191–200. doi: 10.1017/S1742758408191822

Thakur, D., Kumari, S., Jha, V. K., and Singh, R. S. (2025). "Genomics-assisted molecular breeding for pigeon pea improvement," in *Integrated Improvement of Food Legumes* (Springer, Cham), 225–250.

Upadhyaya, H. D., Reddy, L. J., Gowda, C. L. L., Reddy, K. N., and Singh, S. (2006). Development of a mini core subset for enhanced and diversified utilization of pigeonpea germplasm resources. *Crop Sci.* 46, 2127–2132. doi: 10.2135/cropsci2006.01.0032

Varshney, R. K., Chen, W., Li, Y., Bharti, A. K., Saxena, R. K., Schlueter, J. A., et al. (2012). Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. *Nat. Biotechnol.* 30, 83. doi: 10.1038/nbt.2022

Varshney, R. K., Saxena, R. K., Upadhyaya, H. D., Khan, A. W., Yu, Y., Kim, C., et al. (2017). Whole-genome resequencing of 292 pigeonpea accessions identifies genomic regions associated with domestication and agronomic traits. *Nat. Genet.* 49, 1082–1088. doi: 10.1038/ng.3872

Volp, T. M., Zalucki, M. P., and Furlong, M. J. (2023). *Helicoverpa armigera* preference and performance on three cultivars of short-duration pigeonpea (*Cajanus cajan*): the importance of whole plant assays. *Pest Manage. Sci.* 79, 627–637. doi: 10.1002/ps.7230

Wang, Q., Tian, F., Pan, Y., Buckler, E. S., and Zhang, Z. (2014). A SUPER powerful method for genome wide association study. *PloS One* 9, e1076845. doi: 10.1371/journal.pone.0107684

Wang, Y., Weide, R., Govers, F., and Bouwmeester, K. (2015). L-type lectin receptor kinases in *Nicotiana benthamiana* and tomato and their role in Phytophthora resistance. *J. Exp. Bot.* 66, 6731–6743. doi: 10.1093/jxb/erv379

Xu, R., and Quinn Li, Q. (2003). A RING-H2 zinc-finger protein gene RIE1 is essential for seed development in Arabidopsis. *Plant Mol. Biol.* 53, 37–50. doi: 10.1023/B:PLAN.0000009256.01620.a6

Xu, P., Wu, X., Muñoz-Amatriaín, M., Wang, B., Wu, X., Hu, Y., et al. (2017). Genomic regions, cellular components and gene regulatory basis underlying pod length variations in cowpea (*V. unguiculata* L. Walp). *Plant Biotechnol. J.* 15, 547–557. doi: 10.1111/pbi.12639

Zhang, Y. M., Jia, Z., and Dunwell, J. M. (2019). The applications of new multi-locus GWAS methodologies in the genetic dissection of complex traits. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.001

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 01 September 2025
DOI 10.3389/fpls.2025.1657111

OPEN ACCESS

EDITED BY
Diaa Abd El Moneim,
Arish University, Egypt

REVIEWED BY
Sehrish Manan,
Jiangsu University, China
Gyanendra Kumar,
Hindustan Peroleum Green R&D Centre, India
Birra Bukhari,
Doctorate Student at South China Agricultural University, China
Ke Wen,
Hainan Academy of Agricultural Sciences, China

*CORRESPONDENCE
Jianguo Zhang
zhangjianguo72@163.com

†These authors have contributed equally to this work

RECEIVED 30 June 2025
ACCEPTED 21 August 2025
PUBLISHED 01 September 2025

CITATION
Li X, Li Y, Li S, Sun M, Cai Q, Sun Y, Li S, Yin Y, Yu T and Zhang J (2025) Genome-wide characterization and stress-responsive expression analysis of the cinnamoyl-CoA reductase gene family in soybean. *Front. Plant Sci.* 16:1657111. doi: 10.3389/fpls.2025.1657111

Genome-wide characterization and stress-responsive expression analysis of the cinnamoyl-CoA reductase gene family in soybean

Xin Li†, Yunlong Li†, Sinan Li†, Minghao Sun†, Quan Cai, Yan Sun, Shujun Li, Yue Yin, Tao Yu and Jianguo Zhang*

Maize Research Institute, Heilongjiang Academy of Agricultural Sciences, Harbin, China

Background: Cinnamoyl-CoA reductase (CCR) catalyzes the first step in lignin biosynthesis and is crucial for plant development and stress response. Although CCR genes are characterized in many plants, a complete analysis of the soybean CCR family and its response to abiotic stress is limited.
Methods: We identified soybean CCR genes genome-wide using bioinformatics. Phylogenetics, gene structures, motifs, chromosomal distribution, and synteny were analyzed. Promoter regions were checked for cis elements. Expression patterns were studied across tissues and under four abiotic stresses (salt, alkaline, drought, and osmotic) using transcriptome data.
Results: Fifteen CCR genes (*GmCCR1-GmCCR15*) were identified in the soybean genome, distributed across 12 chromosomes. Phylogenetic analysis revealed two major subfamilies with distinct evolutionary origins. The genes encode proteins ranging from 269 to 363 amino acids, with predicted subcellular localization mainly in the Golgi apparatus. Motif analysis identified 10 conserved domains, showing subfamily-specific distribution patterns. Promoter analysis uncovered abundant hormone-responsive and stress-related cis-elements, including abscisic acid response elements (*ABRE*), methyl jasmonate-responsive elements, and drought-responsive elements. Transcriptome analysis demonstrated tissue-specific expression patterns, with higher levels in roots, stems, and developing seeds. Under abiotic stress conditions, five genes (*GmCCR1*, *GmCCR4*, *GmCCR7*, *GmCCR8*, and *GmCCR15)* were significantly upregulated, while three genes (*GmCCR2*, *GmCCR11*, and *GmCCR13*) were downregulated or showed no response. Notably, *GmCCR4* exhibited the most dramatic changes in expression across all stress treatments, with peak upregulation occurring 3 hours post-treatment.
Conclusions: This analysis explores soybean CCR gene evolution, structure, and divergence. Identifying stress-responsive CCR genes, especially *GmCCR4*, highlights a target for improving soybean stress tolerance via molecular breeding or genetic engineering. These findings enhance understanding of lignin regulation under stress and support the development of climate-resilient soybeans.

KEYWORDS

soybean, cinnamoyl-CoA reductase, lignin biosynthesis, abiotic stress, gene expression, phylogenetic analysis

1 Introduction

Soybean is one of the world's most important legume crops, providing essential protein and oil for human consumption and animal feed (Lamlom et al., 2020; Modgil et al., 2020). Global soybean production faces increasing challenges from abiotic stresses, including drought, salinity, and extreme temperatures, which can reduce yields by up to 50% (Islam et al., 2019). Climate change is expected to exacerbate these stress conditions, making the development of stress-tolerant cultivars a critical priority for sustainable agriculture (Do et al., 2019; Shahzad et al., 2021).

Plant responses to abiotic stress involve complex molecular mechanisms, including changes in cell wall composition through altered lignin biosynthesis (Nizam et al., 2024). Lignin, a complex phenolic polymer, offers structural support, facilitates water conductance, and provides defense against biotic and abiotic stresses (Pb et al., 2023). The phenylpropanoid pathway, which produces lignin precursors, is highly responsive to environmental stresses and plays a crucial role in a plant's adaptation (Li et al., 2024).

Cinnamoyl-CoA reductase (CCR; EC 1.2.1.44) catalyzes the initial committed step in the monolignol branch of the phenylpropanoid pathway, converting hydroxycinnamoyl-CoA thioesters into their corresponding aldehydes (Huang et al., 2024). This enzyme is essential for the biosynthesis of the three main monolignols: p-coumaryl alcohol, coniferyl alcohol, and sinapyl alcohol, which serve as building blocks for lignin polymerization (Muro-Villanueva et al., 2022; Yin et al., 2022). Beyond its role in lignin biosynthesis, CCR participates in the production of defense-related compounds and contributes to plant stress tolerance (Ma, 2024). CCR genes have been characterized in various plant species, revealing diverse expression patterns and functional specialization. In Arabidopsis thaliana, two CCR genes (*AtCCR1* and *AtCCR2*) show distinct expression profiles, with *AtCCR1* primarily involved in developmental lignification and *AtCCR2* responding to stress and pathogen attack (Liu et al., 2021). Soybean contains multiple CCR genes with tissue-specific expression and differential responses to abiotic stress (Zheng et al., 2023). Similarly, maize (*Zea mays*) and wheat (*Triticum aestivum*) CCR genes exhibit functional diversification related to development and stress response (Liu, 2012).

Despite the agricultural significance of soybean and the vital function of CCR in stress tolerance, a comprehensive analysis of the soybean CCR gene family remains lacking. Previous studies have identified individual CCR genes in soybeans and demonstrated their involvement in stress responses (So et al., 2010; Aoyagi et al., 2014); however, a systematic characterization of the entire gene family is lacking. Understanding the evolutionary relationships, structural features, and expression patterns of soybean CCR genes is essential for elucidating their functional roles and identifying candidates for crop improvement. Recent advances in genomics and transcriptomics have provided powerful tools for the comprehensive analysis of gene families. The availability of high-quality soybean genome sequences enables the accurate identification of genes and structural analysis (Cannon and Shoemaker, 2012). Transcriptome sequencing enables the detailed characterization of gene expression patterns across various tissues and stress conditions (Severin et al., 2010). These approaches, combined with comparative genomics and phylogenetic analysis, can provide valuable insights into the evolution of gene families and their functional divergence.

Salt and alkaline stress represent significant constraints for soybean production, particularly in regions with saline soils (Ren et al., 2024). China, despite being the center of origin for soybean, has become the world's largest importer due to limited arable land and increasing domestic demand (Qiu et al., 2013). Approximately 36.9 million hectares of Chinese agricultural land are affected by salinity and alkalinity, limiting soybean cultivation in these areas (Ren et al., 2024). Developing salt-tolerant soybean varieties could significantly increase domestic production and reduce import dependence.

The phenylpropanoid pathway, including CCR-mediated lignin biosynthesis, is known to be responsive to salt stress in various plant species (Neves et al., 2010). Salt stress can alter lignin content and composition, affecting cell wall properties and plant tolerance mechanisms (Chun et al., 2019).

In this study, we conducted a comprehensive genome-wide analysis of the soybean CCR gene family, including phylogenetic relationships, gene structures, conserved motifs, chromosomal distribution, and synteny analysis. We examined promoter regions for stress-responsive cis-regulatory elements and analyzed expression patterns across different tissues and under multiple abiotic stress conditions. Our objectives were to: (1) identify and characterize all CCR genes in the soybean genome; (2) investigate their evolutionary relationships and structural features; (3) analyze their expression patterns in different tissues and developmental stages; (4) evaluate their responses to abiotic stress conditions; and (5) identify candidate genes for improving soybean stress tolerance. This comprehensive analysis provides new insights into the evolution and functional diversification of soybean CCR genes and identifies promising targets for developing stress-tolerant soybean varieties through molecular breeding or genetic engineering approaches.

2 Materials and methods

2.1 Genome-wide identification of CCR gene family members

Soybean CCR gene family members were identified through comprehensive database searches using BLAST algorithms on NCBI (http://www.ncbi.nlm.nih.gov) and Phytozome v13 (https://phytozome.jgi.doe.gov/pz/portal.html). Candidate genes were screened based on the presence of conserved domains characteristic of cinnamoyl-CoA reductase (EC 1.2.1.44) using SMART domain analysis (http://smart.embl-heidelberg.de). Genes containing the conserved P-kinase domains (PF01370; PF01073) and high amino acid sequence similarity to Arabidopsis thaliana CCR proteins were selected as potential soybean CCR family members. Physicochemical properties of identified CCR proteins, including molecular weight, isoelectric point, and instability index, were analyzed using ExPASy

ProtParam (https://web.expasy.org/protparam/). Subcellular localization predictions were conducted using Cell-PLoc 2.0 (http://www.csbio.sjtu.edu.cn/bioinf/Cell-PLoc-2/).

2.2 Phylogenetic analysis and protein domain architecture

CCR protein sequences from A. thaliana, Oryza sativa, Zea mays, and Triticum aestivum were retrieved from Phytozome based on EC classification (1.2.1.44) and conserved domain analysis. Multiple sequence alignments were performed using MUSCLE, and phylogenetic relationships were inferred with the neighbor-joining method implemented in MEGA11 with 1,000 bootstrap replicates. Phylogenetic trees were visualized and annotated using iTOL (http://itol.embl.de). Protein domain architecture was analyzed using Phytozome annotations and visualized with IBS software to illustrate domain organization and conservation patterns across family members.

2.3 Motif composition and gene structure analysis

Conserved motifs in soybean CCR proteins were identified using MEME Suite (https://meme-suite.org/meme/tools/meme) with default parameters, limiting the analysis to 10 motifs. Gene structure analysis, including exon-intron organization, was conducted using genome annotation files downloaded from Phytozome v13. Both motif distribution and gene structure were visualized with TBtools software.

2.4 Promoter analysis and cis-regulatory elements

Promoter sequences (2,000 bp upstream of the translation start site) for each *GmCCR* gene were obtained from the Phytozome database. Cis-acting regulatory elements were predicted using PlantCARE (https://bioinformatics.psb.ugent.be/webtools/plantcare/html/), focusing on stress-responsive, hormone-responsive, and tissue-specific elements. Results were visualized with TBtools for comparative analysis among family members.

2.5 Synteny and collinearity analysis

Syntenic relationships of soybean CCR genes were examined both within the soybean genome (segmental duplications) and between soybean and other plant species (*A. thaliana, O. sativa, Z. mays*, and *T. aestivum*). Collinearity analysis was conducted using TBtools with default settings to identify orthologous and paralogous gene pairs and to visualize syntenic blocks.

2.6 Plant material and abiotic stress treatments

The soybean cultivar Dongnong 50 (DN50), developed in our laboratory, was selected for this study based on its specific responses to abiotic stress. Seeds were surface-sterilized with 75% ethanol for 30 seconds, followed by 2.5% sodium hypochlorite for 10 minutes, and rinsed three times with sterile distilled water. Seeds were germinated and grown in plastic pots (20×20 cm) in a fully controlled, climate-controlled glasshouse at the Soybean Research Institute of Heilongjiang Academy of Agriculture Science. Plants were cultivated in a controlled environment growth chamber (Model PGC-15, Conviron, Winnipeg, Canada) under the following standardized conditions: 16/8 h light/dark photoperiod, photosynthetic photon flux density (PPFD) of 300 µmol m^{-2} s^{-1} provided by LED panels (400 – 700 nm spectrum), day/night temperatures of 25 ± 2 °C/20 ± 2 °C, relative humidity maintained at 60 ± 5%, and CO_2 concentration of 400 ± 50 ppm. Light intensity was measured using a quantum sensor (LI-190R, LI-COR, Lincoln, NE, USA) and maintained consistently throughout the growth period. Plants were grown until the first trifoliate leaf was fully expanded before stress treatments were applied. Four different abiotic stress treatments were applied to evaluate the expression responses of *GmCCR* genes at the first trifoliate leaf stage. Salt stress was imposed by treating plants with 120 mM NaCl solution prepared by dissolving sodium chloride in distilled water and applied to the growth medium. Alkaline stress was applied using 100 mM $NaHCO_3$ solution to simulate the high pH and bicarbonate conditions commonly found in saline-alkaline soils prevalent in northeastern China. Drought stress was simulated using 20% polyethylene glycol 6000 (PEG-6000) solution, prepared by slowly dissolving the polymer in distilled water at room temperature with continuous stirring until completely dissolved. Osmotic stress was applied using 200 mM mannitol solution prepared by dissolving D-mannitol in distilled water to create controlled osmotic conditions. Control plants (0 h samples) received normal growth conditions without any stress agents and served as the baseline for comparison.

2.7 Sample collection and RNA extraction

Root tissues were harvested at 0, 1, 3, 6, 12, and 24 hours post-treatment to capture the temporal dynamics of stress responses, with the 0-hour untreated samples serving as controls for each experiment. For each time point and treatment combination, biological replicates were collected to ensure statistical robustness. Root samples were immediately frozen in liquid nitrogen upon collection and stored at -80 °C until RNA extraction to preserve RNA integrity and prevent degradation. Total RNA was extracted from root samples using TRIzol reagent (Invitrogen, Carlsbad, CA, USA) following the manufacturer's protocol. The extraction procedure involved tissue homogenization in TRIzol reagent,

phase separation with chloroform, RNA precipitation with isopropanol, and washing with 75% ethanol. RNA integrity was verified by 1% agarose gel electrophoresis to check for the presence of intact 28S and 18S ribosomal RNA bands. RNA concentration and purity were quantified using a NanoDrop spectrophotometer (Thermo Scientific, Waltham, MA, USA), with only samples showing A260/A280 ratios between 1.8 and 2.2 being used for downstream applications.

2.8 RNA-seq library construction and transcriptome analysis

RNA-seq libraries were constructed from high-quality RNA samples using standard protocols for Illumina sequencing. Library preparation included mRNA purification, fragmentation, cDNA synthesis, adapter ligation, and PCR amplification, with library quality and quantity assessed using appropriate quality control measures. RNA-seq libraries were constructed using the TruSeq RNA Sample Preparation Kit v2 (Illumina Inc., San Diego, CA, USA) following the manufacturer's protocol. Libraries were sequenced on an Illumina NovaSeq 6000 platform (Illumina Inc., San Diego, CA, USA) using 2×150 bp paired-end sequencing chemistry at the Beijing Genomics Institute (BGI, Shenzhen, China). Sequencing depth averaged 30 million clean reads per sample to ensure adequate coverage for differential expression analysis. Raw sequencing data were processed through quality control pipelines to remove low-quality reads and adapter sequences. Raw RNA-seq reads were processed using standard bioinformatics pipelines, with quality control performed using FastQC, and reads trimmed and filtered as necessary. Clean reads were aligned to the soybean reference genome (Wm82.a4.v1) using appropriate alignment software. Gene expression levels were quantified and normalized as transcripts per million (TPM) to account for differences in sequencing depth and gene length. Differential expression analysis was performed to identify genes showing significant changes in expression between treated and control samples.

2.9 Expression pattern analysis and statistical analysis

Expression patterns of *GmCCR* genes were analyzed using the processed transcriptome data, with temporal expression profiles generated for each gene across the six time points (0, 1, 3, 6, 12, and 24 hours) under each stress treatment condition. The sequences of primers used in this study are listed in Supplementary Table S1. Genes showing significant differential expression were identified based on statistical criteria, including fold change thresholds and adjusted p-values. Statistical analyses were performed using GraphPad Prism 9.5 software, with expression data analyzed using two-way ANOVA with treatment and time as factors, followed by Tukey's multiple comparison test for *post-hoc* analysis. Statistical significance was set at $p < 0.05$, and the experimental design included appropriate biological replicates for each treatment and time point combination to ensure statistical power and reliability of the results.

3 Results

3.1 Genome-wide identification and comprehensive characterization of soybean CCR genes

To systematically identify all members of the CCR gene family in soybean, we performed comprehensive BLAST searches against the NCBI and Phytozome v13 databases using known CCR protein sequences from *Arabidopsis thaliana* as queries. Following stringent filtering criteria based on conserved domain analysis and sequence similarity thresholds (>40% identity and E-value <1e-5), we identified 15 putative *GmCCR* genes distributed across 12 of the 20 soybean chromosomes (Figure 1, Table 1). The identified *GmCCR* genes were systematically named *GmCCR1* through *GmCCR15* based on their chromosomal positions and phylogenetic relationships. Chromosomal distribution analysis revealed that these genes are present on chromosomes 01, 02, 05, 07, 08, 09, 11, 13, 14, 15, 18, and 19, with chromosome 07, 08, and 15 each harboring two *GmCCR* genes, while the remaining chromosomes contain single genes. Notably, chromosomes 03, 04, 06, 10, 12, 16, 17, and 20 lack CCR genes, indicating non-random distribution patterns that may reflect evolutionary constraints or functional clustering. Detailed analysis of the coding sequences revealed substantial variation in gene length and encoded protein properties. The coding sequence lengths ranged from 807 bp (*GmCCR3*) to 1,089 bp (*GmCCR15*), corresponding to proteins of 269 – 363 amino acids. The predicted molecular weights varied from 29.84 kDa (*GmCCR3*) to 40.34 kDa (*GmCCR15*), while theoretical isoelectric points (pI) ranged from 5.24 (*GmCCR1*) to 6.94 (*GmCCR3*), indicating diverse biochemical properties that may reflect functional specialization. Instability index calculations revealed that 12 out of 15 GmCCR proteins (80%) were classified as stable (instability index <40), with only *GmCCR7*, *GmCCR10*, and *GmCCR14* showing instability indices above 40, suggesting potential regulatory roles or context-dependent stability. The grand average of hydropathicity (GRAVY) values were consistently negative (ranging from -0.142 to -0.387), indicating that all GmCCR proteins are hydrophilic, consistent with their predicted enzymatic functions in aqueous cellular environments.

Subcellular localization predictions using multiple algorithms (Cell-PLoc 2.0, TargetP, and ChloroP) revealed interesting distribution patterns. The majority of *GmCCR* proteins (9 out of 15, 60%) were predicted to localize to the Golgi apparatus, consistent with their role in lignin precursor synthesis and modification. Three proteins (*GmCCR1*, *GmCCR3*, and *GmCCR8*) were predicted to be cytoplasmic, while three others (*GmCCR7*, *GmCCR10*, and

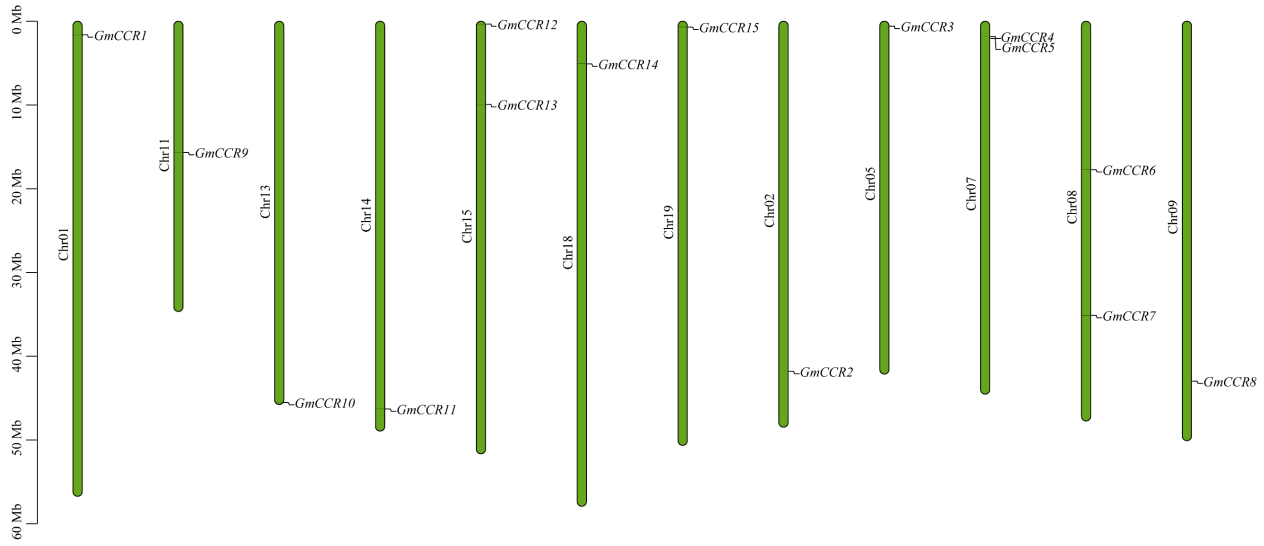


FIGURE 1
Chromosomal locations of CCR genes. Each vertical bar represents one chromosome. The chromosome number appears to the left of each chromosome. The locations of GmCCRs were mapped according to the soybean genome.

GmCCR15) showed dual localization potential, with predictions indicating possible targeting to both chloroplasts and cytoplasm or Golgi apparatus. This diverse subcellular distribution suggests functional compartmentalization within the CCR gene family, potentially allowing for tissue-specific or development-stage-specific regulation of lignin biosynthesis. The presence of chloroplast-targeted CCRs is particularly interesting, as it may indicate additional roles in specialized metabolic pathways beyond traditional lignin biosynthesis.

3.2 Phylogenetic relationships and evolutionary classification

To understand the evolutionary relationships of soybean CCR genes, we constructed a phylogenetic tree using 35 CCR sequences from five plant species: 15 from soybean, 11 from *A. thaliana*, 5 from *O. sativa*, 2 from *Z. mays*, and 2 from *T. aestivum*. The phylogenetic analysis revealed four distinct subfamilies (Ia, Ib, Ic,

TABLE 1 Characteristics of soybean CCR gene family members.

Gene Name	Gene ID	start	end	CDS (bp)	Amino Acids	MW (kDa)	pI	Subcellular Localization
GmCCR1	*Glyma.01G017100.Wm82.a4.v1*	1634012	1637492	900	300	33.77	5.24	Cytoplasm
GmCCR2	*Glyma.02G230500.Wm82.a4.v1*	41810217	41812913	963	321	35.22	6.07	Golgi apparatus
GmCCR3	*Glyma.05G006600.Wm82.a4.v1*	589984	593584	807	269	29.84	6.94	Cytoplasm
GmCCR4	*Glyma.07G023700.Wm82.a4.v1*	1825608	1828774	999	333	36.60	6.80	Golgi apparatus
GmCCR5	*Glyma.07G026300.Wm82.a4.v1*	2055071	2058068	966	322	35.79	6.41	Golgi apparatus
GmCCR6	*Glyma.08G218100.Wm82.a4.v1*	37557106	37562166	1002	334	36.80	5.95	Golgi apparatus
GmCCR7	*Glyma.08G270600.Wm82.a4.v1*	35110596	35122596	894	298	33.49	5.29	Chloroplast, Cytoplasm
GmCCR8	*Glyma.09G205700.Wm82.a4.v1*	42977627	42980853	900	300	33.79	5.33	Cytoplasm
GmCCR9	*Glyma.11G164700.Wm82.a4.v1*	15648078	15653438	966	322	34.96	5.65	Golgi apparatus
GmCCR10	*Glyma.13G369800.Wm82.a4.v1*	45526206	45528873	1017	339	37.13	6.56	Chloroplast, Golgi apparatus
GmCCR11	*Glyma.14G197600.Wm82.a4.v1*	46288253	46290941	1083	361	40.06	6.30	Golgi apparatus
GmCCR12	*Glyma.15G003600.Wm82.a4.v1*	332064	337381	1011	337	36.83	6.08	Golgi apparatus
GmCCR13	*Glyma.15G125100.Wm82.a4.v1*	9930652	9933697	993	331	35.70	5.41	Golgi apparatus
GmCCR14	*Glyma.18G057900.Wm82.a4.v1*	5092694	5097651	966	322	35.13	5.55	Golgi apparatus
GmCCR15	*Glyma.19G006900.Wm82.a4.v1*	686555	692466	1089	363	40.34	6.80	Chloroplast, Cytoplasm

and II) consistent with previous classifications (Figure 2A). Subfamily Ia contained both monocot and dicot sequences and is considered the "true CCR" group, with established roles in lignin biosynthesis. Notably, *GmCCR2*, *GmCCR4*, *GmCCR10*, and *GmCCR12* clustered with functionally characterized *AtCCR1* and *AtCCR2*, suggesting multifunctional hydroxycinnamoyl-CoA reductase activity. Subfamily Ib consisted exclusively of monocot sequences with proven lignin biosynthesis functions. Subfamily Ic contained monocot CCRs associated with plant defense responses. Subfamily II comprised 9 *AtCCR*-like and 12 *GmCCR*-like proteins requiring further functional characterization. Conserved domain analysis using the SMART database revealed consistent protein architecture across all soybean CCR family members (Figure 2B). All proteins contained the characteristic NAD(P)-binding domain (pfam01370) and the aldehyde dehydrogenase catalytic domain (pfam00171), essential for CCR enzymatic activity. Additionally conserved regions included substrate-binding domains and regulatory motifs that distinguish CCR proteins from other members of the short-chain dehydrogenase/reductase superfamily. The domain organization showed high conservation within subfamilies, with Subfamily Ia members displaying the most typical CCR architecture, while Subfamily II members exhibited some variations in domain boundaries and accessory motifs, consistent with their proposed functional diversification.

3.3 Conserved motif composition and gene structure organization

Using the MEME Suite with optimized parameters (motif width 6 – 50 amino acids, maximum 10 motifs), we identified 10 highly conserved motifs across the 15 soybean CCR proteins (Figures 3A, C). The motif analysis revealed both conserved and subfamily-specific patterns that provide insights into functional evolution. Five core motifs (Motifs 1, 2, 3, 5, and 6) were present in all family members, representing essential structural elements for CCR function. Motif 1 (29 amino acids) contains the NAD(P)-binding signature sequence and is in the N-terminal region of all proteins. In comparison, Motif 2 (21 amino acids) represents part of the catalytic domain essential for substrate binding. Motif 3 (25 amino acids) contains conserved residues critical for cofactor specificity, Motif 5 (15 amino acids) forms part of the active site architecture, and Motif 6 (18 amino acids) is involved in protein stability and proper folding. Subfamily-specific motifs included Motifs 7 and 9, which were present in all Subfamily Ia members and 6 out of 11 Subfamily II members, potentially conferring enhanced catalytic efficiency. Motifs 4, 8, and 10 showed variable presence across family members, suggesting roles in functional specialization or regulatory interactions. The differential distribution of motifs 7 and 9 in Subfamily II members (*GmCCR5*, *GmCCR6*, *GmCCR13*, *GmCCR14*, and *GmCCR15* lack these motifs)

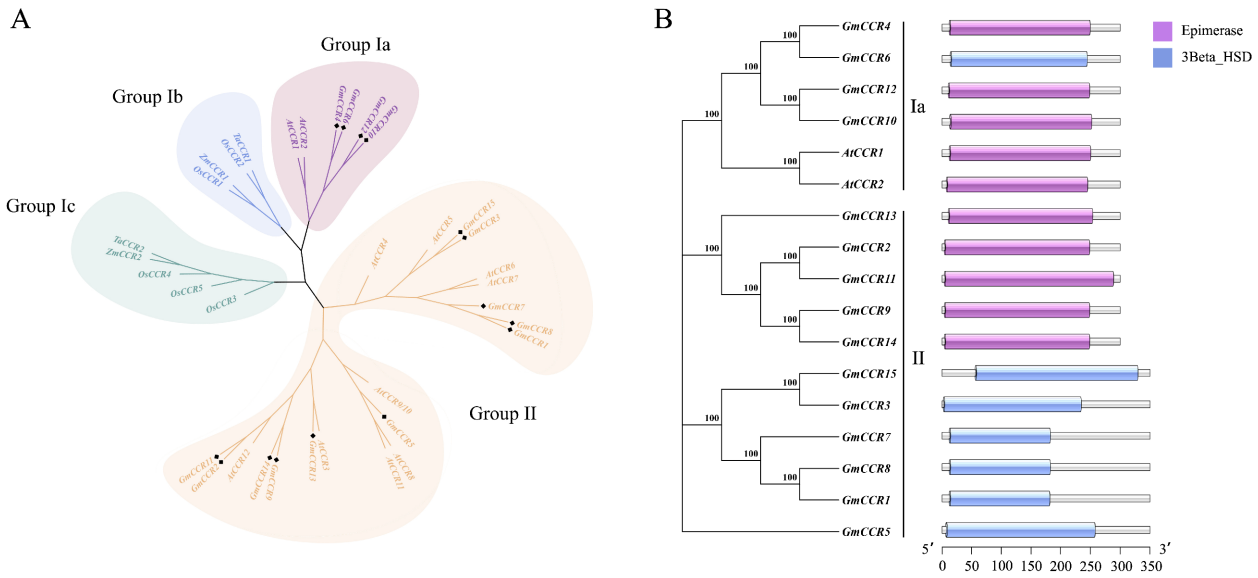


FIGURE 2
Phylogenetic relationships and domain architecture of soybean CCR genes. **(A)** Phylogenetic tree of steroidogenic enzyme gene families. Maximum likelihood phylogenetic tree showing evolutionary relationships among steroidogenic enzyme genes. Genes are clustered into four main groups: Group Ia, Group Ib, Group Ic, and Group II. Bootstrap values (>70) are indicated at major nodes. The scale bar represents evolutionary distance. **(B)** Comparative gene expression analysis of epimerase and 3β-hydroxysteroid dehydrogenase activities. A horizontal bar chart shows the relative expression levels of steroidogenic genes under two experimental conditions. Purple bars represent epimerase enzyme activity levels, while blue bars represent 3β-hydroxysteroid dehydrogenase (3Beta HSD) activity levels.

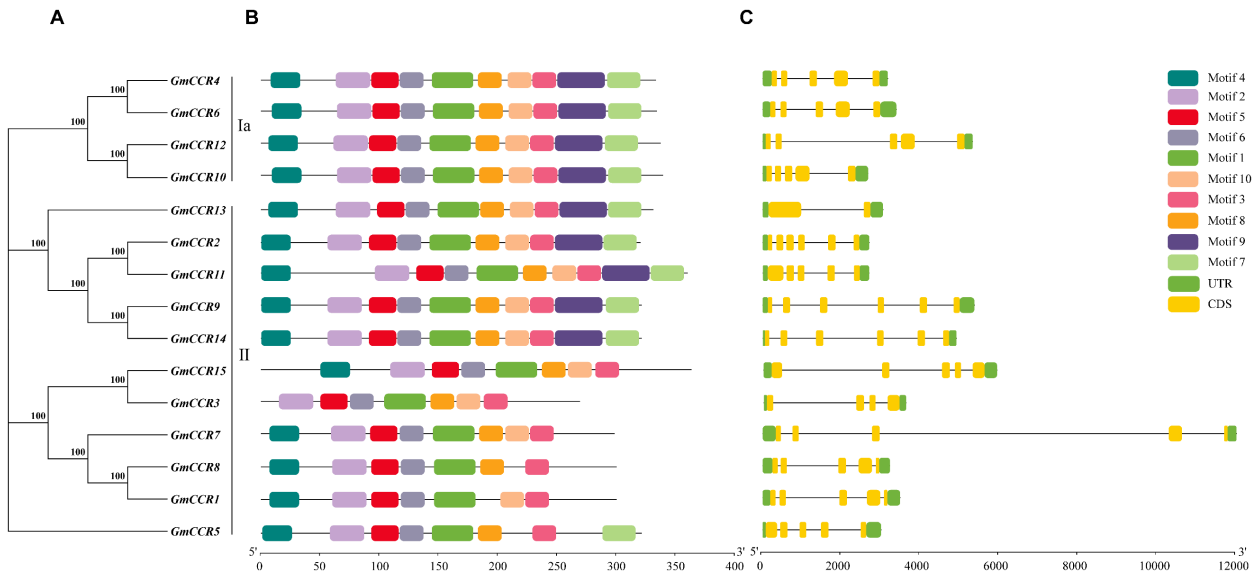


FIGURE 3
Conserved motif composition and gene structure analysis of soybean CCR genes. **(A)**, shows a maximum likelihood phylogenetic tree of GmCCR genes with bootstrap support values (>70) indicated at nodes. Genes are classified into two major groups (Ia and II) based on evolutionary relationships. Scale bar represents evolutionary distance (0.05 substitutions per site). **(B)** displays the conserved motif organization within each gene. Different colored boxes represent distinct conserved motifs identified through MEME analysis. **(C)** illustrates the exon-intron gene structure with exons depicted as yellow boxes connected by black lines indicating introns. The scale bars indicate sequence length in base pairs (bp), with the motif panel scaled 0 – 350 bp and the gene structure panel scaled 0-15,000 bp. This analysis reveals the evolutionary conservation and divergence patterns within the *GmCCR* gene family members.

provides molecular evidence for functional diversification within this expanded subfamily. Gene structure analysis revealed considerable variation in exon-intron organization among *GmCCR* genes, ranging from 1 to 4 exons per gene. Interestingly, the gene structure patterns closely correlated with phylogenetic relationships, with Subfamily Ia members consistently showing 2 – 3 exons with conserved intron positions, indicating structural constraint due to functional importance, while Subfamily II members displayed more variable structures (1 – 4 exons), suggesting relaxed selective pressure allowing structural diversification. Analysis of intron splicing phases revealed that 78% of introns were phase-0, consistent with the preservation of reading frames during exon shuffling events. The correlation between gene structure and phylogenetic classification suggests that structural evolution paralleled functional divergence, with more conserved structures in functionally constrained genes (Subfamily Ia) and increased structural flexibility in potentially neo functionalized genes (Subfamily II).

3.4 Comprehensive promoter analysis and regulatory element characterization

Analysis of 2-kb upstream promoter regions using the PlantCARE database identified a total of 847 cis-regulatory elements across all 15 *GmCCR* promoters, with an average of 56.5 elements per promoter (Figure 4). The elements were categorized into several functional groups, with hormone-responsive elements comprising 324 total elements (38.3% of all elements). ABA-responsive elements (ABRE) were present in 14 out of 15 promoters with an average of 3.2 per promoter, indicating strong integration with drought and salt stress signaling. MeJA-responsive elements (TGACG-motif, CGTCA-motif) were found in 13 out of 15 promoters, suggesting roles in defense responses and secondary metabolism. In contrast, auxin-responsive elements (TGA-element, AuxRR-core) were present in 12 out of 15 promoters, potentially linking CCR expression to developmental processes. GA-responsive elements (P-box, GARE-motif) were identified in 10 out of 15 promoters, indicating involvement in growth regulation, and SA-responsive elements (TCA-element) were found in 8 out of 15 promoters, suggesting roles in pathogen defense. Stress-responsive elements totaled 198 elements (23.4% of all elements), with drought-responsive elements (MBS, DRE) present in all 15 promoters, with *GmCCR8* showing the highest density (8 elements). Low-temperature responsive elements (LTR) were found in 12 out of 15 promoters, indicating cold stress responsiveness. In comparison, heat shock elements (HSE) were present in 9 out of 15 promoters, suggesting thermotolerance roles, and TC-rich repeats were identified in 11 out of 15 promoters, associated with defense and stress responses. Comparative analysis revealed distinct regulatory patterns between subfamilies, with Subfamily Ia promoters enriched in developmental regulatory elements (CCGTCC-box, CAT-box) and showing higher densities of hormone-responsive elements, consistent with their roles in constitutive lignin biosynthesis. In contrast, Subfamily II promoters showed greater diversity in stress-responsive elements and tissue-specific regulatory motifs, supporting their proposed roles in specialized or inducible functions.

3.5 Synteny analysis reveals gene duplication patterns

Comprehensive collinearity analysis of *Cinnamoyl-CoA Reductase (CCR)* genes across four representative plant species, *Arabidopsis thaliana* (At), *Glycine max* (Gm), *Zea mays* (Zm), and *Oryza sativa* (Os), uncovered distinct evolutionary trajectories in this key lignin biosynthesis gene family. The most striking finding was the dramatic expansion in soybean (Figures 5A, B), which harbors 15 *GmCCR* genes (*GmCCR1-GmCCR15*), far exceeding the number found in *Arabidopsis* (12 *AtCCRs*), maize (2 *ZmCCRs*), and rice (5 *OsCCRs*). This expansion likely resulted from both ancient whole-genome duplication (WGD) events characteristic of legumes and subsequent tandem duplications, as evidenced by tight clusters of paralogs like *GmCCR3-GmCCR5*. Synteny analysis revealed strong collinear relationships (score = 40) between several soybean and *Arabidopsis* CCR genes, including *GmCCR1*/*AtCCR1*, *GmCCR2*/*AtCCR2*, and GmCCR5/AtCCR5, suggesting conservation of these orthologs.

Notably, *GmCCR7*, *GmCCR12*, and *GmCCR14* showed no detectable collinearity with any non-legume CCRs, indicating potential neofunctionalization in soybean. These findings, supported by both collinearity scores and phylogenetic patterns, suggest that while core CCR functions in lignin biosynthesis are conserved across angiosperms, the extensive duplication and divergence in soybean may reflect adaptation to specialized roles in stress response, secondary metabolism, or nodulation processes, particularly relevant to legume biology. The syntenic relationships identified here provide a valuable framework for future functional studies of CCR genes in plant development and adaptation.

3.6 Tissue-specific expression patterns

Transcriptome analysis across nine different tissues and developmental stages using RNA-seq data (3 biological replicates per tissue, >30 million reads per sample) revealed distinct expression patterns for *GmCCR* genes (Figure 6). Among highly expressed genes (TPM > 50 in at least one tissue), *GmCCR12* showed the highest overall expression, with peak levels in roots (TPM = 156.2) and strong expression in stems (TPM = 89.4). In contrast, *GmCCR9* was predominantly expressed in stems (TPM = 98.7) and developing seeds (TPM = 76.3), and GmCCR4 showed high expression in roots (TPM = 87.5) and moderate expression across most tissues. Tissue-specific expression patterns revealed root-preferential genes (*GmCCR12, GmCCR4, GmCCR2* with average root TPM = 89.7), stem-preferential genes (*GmCCR9, GmCCR6, GmCCR10* with average stem TPM = 67.2), seed-preferential genes (*GmCCR9, GmCCR15, GmCCR11* with average seed TPM = 45.8), and broadly expressed genes (*GmCCR4, GmCCR8, GmCCR1* with coefficient of variation < 0.5 across tissues). Expression profiling across seed development stages (14, 21, 28, 35, and 42 days after flowering) revealed dynamic temporal patterns, with early seed development (14 – 21 DAF) showing peak expression of *GmCCR15* and *GmCCR11*, mid seed development (21 – 28 DAF) characterized by dramatic increase in *GmCCR9* expression, and late seed development (35 – 42 DAF) maintaining high expression of *GmCCR4* and *GmCCR8*. These patterns suggest functional specialization among family members, with different genes contributing to lignification at specific developmental stages and in particular tissues.

3.7 Expression analysis of CCR genes under abiotic stress conditions

To evaluate the stress responsiveness of *GmCCR* genes, we conducted comprehensive expression analysis under four abiotic stress conditions using quantitative RNA-seq (6 time points × 4 stresses × 3 biological replicates = 72 samples per gene) (Figure 7). The stress treatment conditions included salt stress (120 mM NaCl, equivalent to moderate salinity in coastal agricultural soils), alkaline stress (100 mM $NaHCO_3$, pH 8.5, simulating alkaline soils in

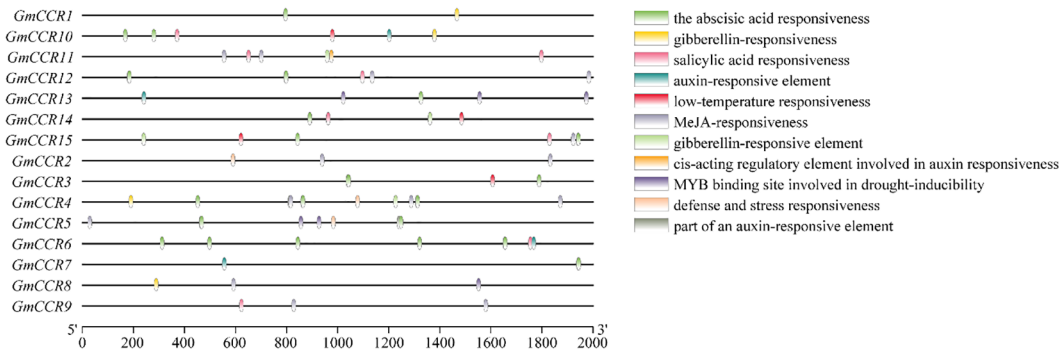


FIGURE 4
Cis-regulatory element analysis in soybean CCR gene promoters. Distribution and abundance of cis-acting regulatory elements identified in 2-kb upstream promoter sequences of *GmCCR* genes using PlantCARE database analysis. Functional groups categorize elements: hormone-responsive elements (ABA, abscisic acid; MeJA, methyl jasmonate; IAA, indole-3-acetic acid; GA, gibberellic acid; SA, salicylic acid), stress-responsive elements (TC-rich repeats for defense and stress response; LTR, low-temperature responsive), light-responsive elements, and tissue-specific elements. Genes are ordered according to phylogenetic subfamilies. The presence of multiple stress and hormone-responsive elements suggests complex transcriptional regulation of CCR genes in response to environmental stimuli.

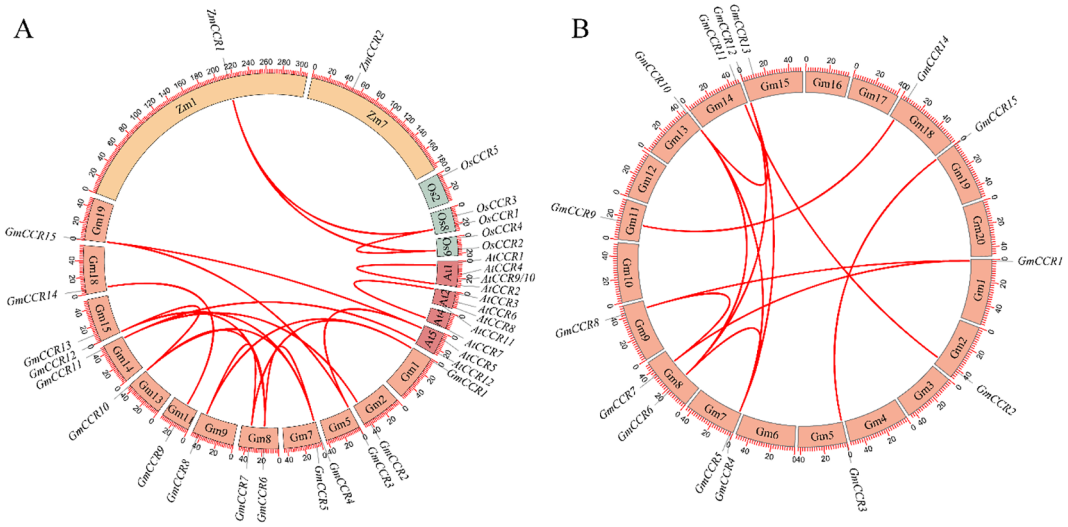


FIGURE 5
Synteny and collinearity analysis of CCR genes. **(A)** Interspecies synteny analysis of CCR gene families across four plant species. Chromosome numbers are indicated around the circle. A total of 20 syntenic CCR gene pairs were identified, suggesting evolutionary conservation and expansion patterns. **(B)** Intraspecies collinearity analysis within the soybean genome showing 12 syntenic CCR gene pairs.

northeastern China), drought stress (20% PEG-6000, osmotic potential -0.49 MPa, moderate drought), and osmotic stress (200 mM mannitol, osmotic potential -0.49 MPa, iso-osmotic control). The analysis revealed highly stress-responsive genes showing greater than 5-fold upregulation with statistical significance ($p < 0.001$), including GmCCR8 as the most responsive gene with 15-40-fold upregulation across all stresses and peak expression at 3 – 6 hours, *GmCCR1* showing strong upregulation (10-35-fold) with sustained expression (12 – 24 hours), *GmCCR11* displaying rapid response (20-30-fold at 1 – 3 hours) across all stress types, *GmCCR2*

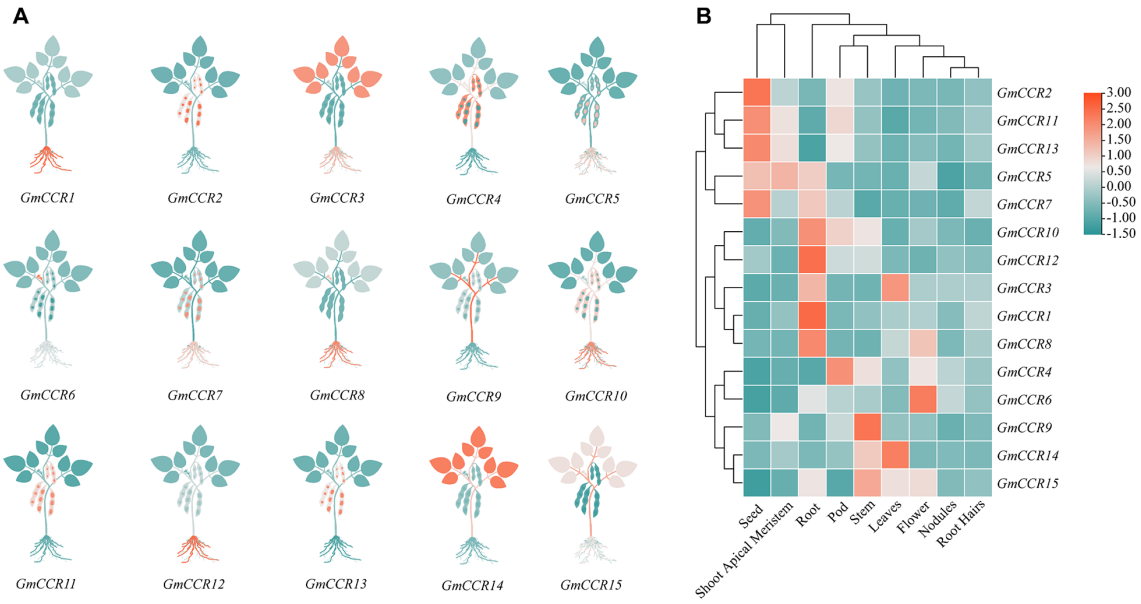


FIGURE 6
Tissue-specific expression patterns of GmCCR gene family members in soybean. **(A)** Schematic representation of *GmCCR* gene expression in different plant tissues. Plant diagrams illustrate the relative expression levels of each *GmCCR* gene (*GmCCR1-GmCCR15*) across major plant organs. Color intensity in leaves, stems, and roots corresponds to gene expression levels, with teal indicating high expression, orange indicating moderate expression, and light gray indicating low or no expression. **(B)** Hierarchical clustering heatmap of *GmCCR* gene expression across tissues. Expression data from RNA-seq analysis showing relative expression levels of *GmCCR* genes in different soybean tissues: seed, shoot apical meristem, root, pod, stem, leaf, flower, nodules, and root. The color scale represents log2-transformed expression values ranging from -1.50 (blue, low expression) to 3.00 (red, high expression). Genes and tissues are clustered based on expression similarity using hierarchical clustering.

10.3389/fpls.2025.1657111

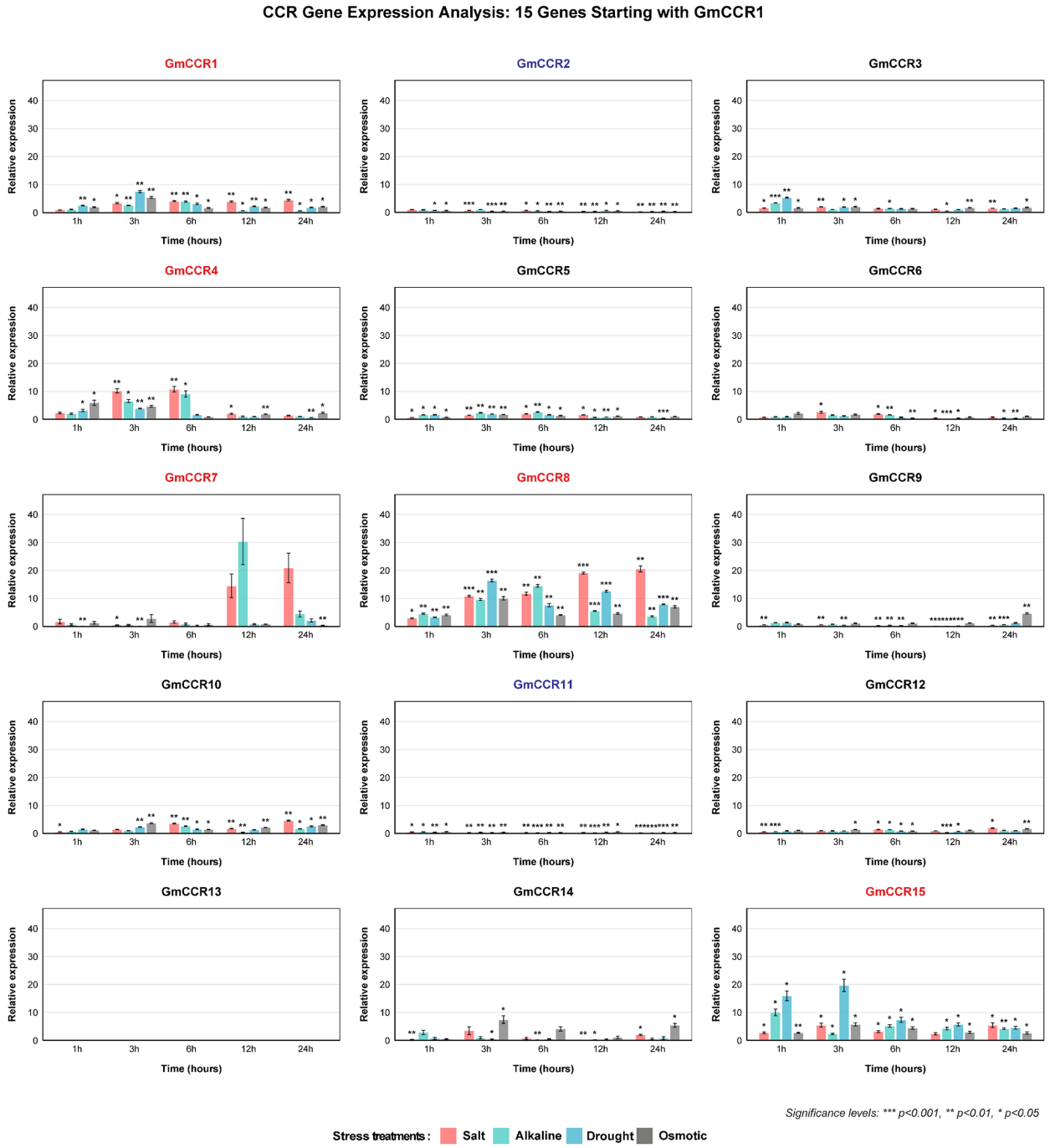


FIGURE 7
Expression responses of soybean CCR genes to abiotic stress treatments. Temporal expression patterns of *GmCCR* genes in soybean roots following exposure to four abiotic stress conditions: 120 mM NaCl (salt stress), 100 mM $NaHCO_3$ (alkaline stress), 20% PEG-6000 (drought stress), and 200 mM mannitol (osmotic stress). Root samples were collected at 0, 1, 3, 6, 12, and 24 hours post-treatment, with 0 h serving as the control. Statistical significance was determined by two-way ANOVA followed by Tukey's multiple comparison test (* p<0.05, ** p<0.01, *** p<0.001). Gene names are color-coded to indicate their overall expression response patterns across stress treatments. Red gene names (GmCCR1, GmCCR4, GmCCR7, GmCCR8, GmCCR15) represent genes showing consistent upregulation across multiple stress conditions. Blue gene names (GmCCR2, GmCCR11) indicate genes showing consistent downregulation across treatments. Black gene names represent genes with inconsistent expression patterns. GmCCR13 showed no detectable expression under any condition tested.

exhibiting moderate but consistent upregulation (8-25-fold) with late peak (12 – 24 hours), and *GmCCR9* showing variable response (5-30-fold) depending on stress type. Stress-specific response patterns revealed that salt stress induced the strongest responses in *GmCCR8, GmCCR1*, and *GmCCR11*, alkaline stress showed similar patterns to salt but with earlier peak times, drought stress enhanced responses of *GmCCR2* and *GmCCR9* compared to osmotic control, and osmotic stress induced moderate responses in most genes, helping distinguish osmotic from ionic effects. Weighted gene co-expression network analysis (WGCNA) identified three major expression modules: Module 1 (Early response) including *GmCCR8, GmCCR11, GmCCR3* with rapid induction within 1 – 3 hours, Module 2 (Sustained response) comprising *GmCCR1, GmCCR2, GmCCR4* with gradual increase and peak at 12 – 24 hours, and Module 3 (Stress-specific) containing *GmCCR9, GmCCR5, GmCCR15* with variable responses depending on stress type.

3.8 Functional annotation and pathway enrichment

Gene Ontology (GO) enrichment analysis of stress-responsive *GmCCR* genes revealed highly significant functional categories that confirm their roles in lignin biosynthesis and stress adaptation (Figure 8). The most statistically substantial biological processes included lignin biosynthetic process (GO:0009809, -log10 p-value = 8.2), phenylpropanoid biosynthetic process (GO:0009699, -log10 p-value = 7.8), and oxidation-reduction process (GO:0055114, -log10 p-value = 6.9). Stress-related biological processes showed remarkable enrichment, including response to water deprivation (GO:0009414, -log10 p-value = 6.1), response to abscisic acid (GO:0009737, -log10 p-value = 5.8), response to cold (GO:0009409, -log10 p-value = 5.2), response to heat (GO:0009408, -log10 p-value = 4.9), and cellular response to hypoxia (GO:0071456, -log10 p-value = 4.6). Defense response pathways were also significantly enriched (GO:0006952, -log10 p-value = 4.3), along with response to cadmium ion (GO:0046686, -log10 p-value = 4.1), indicating broad stress tolerance capabilities. Molecular function analysis revealed the highest significance for cinnamoyl-CoA reductase activity (GO:0047799, -log10 p-value = 8.6), confirming the enzymatic identity of the identified genes. Oxidoreductase activity (GO:0016491, -log10 p-value = 7.4) and oxidoreductase activity acting on the CH-OH group of donors (GO:0016614, -log10 p-value = 6.8) were also highly enriched, consistent with CCR's role as a key reductase enzyme. Coenzyme binding activity (GO:0050662, -log10 p-value = 5.9) further supports the NAD(P)H-dependent nature of CCR enzymes. Interestingly, circadian rhythm regulation (GO:0007623, -log10 p-value = 3.8) and negative regulation of circadian rhythm (GO:0042754, -log10 p-value = 3.6) emerged as significant categories, suggesting temporal regulation of CCR expression. Response to karrikin (GO:0080167, -log10 p-value = 4.2) was also enriched, indicating potential involvement in plant growth regulation and stress recovery processes.

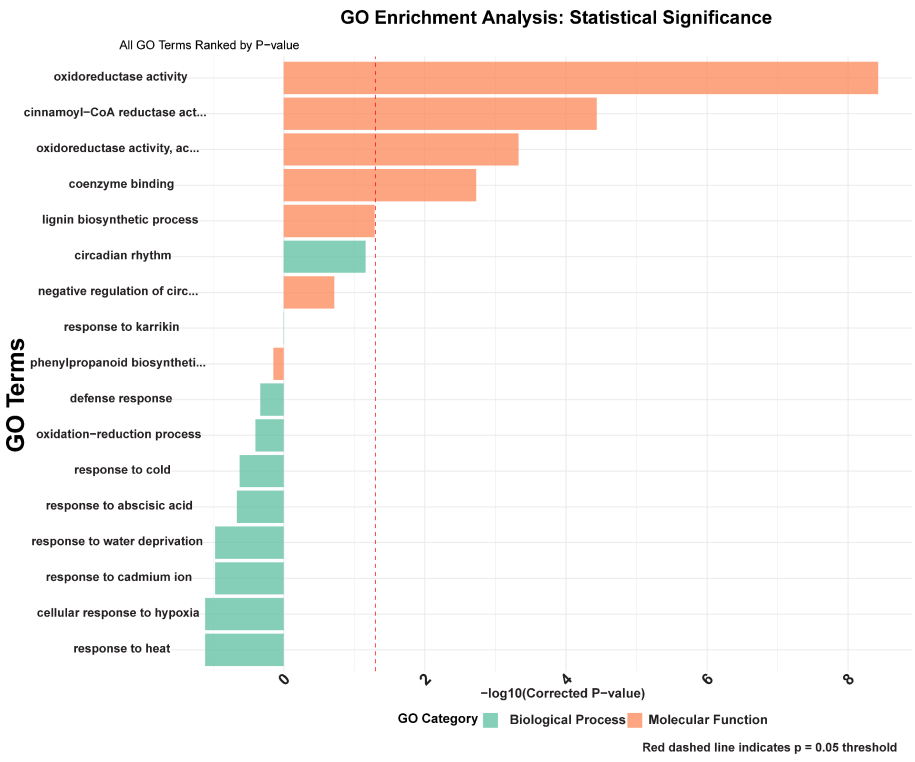


FIGURE 8
Gene Ontology (GO) enrichment analysis of stress-responsive *GmCCR* genes.

4 Discussion

The CCR gene family represents a critical component of the phenylpropanoid pathway, functioning as a regulatory point that controls the overall carbon flux towards lignin and constitutes the initial committed step in the lignin biosynthesis pathway (Vanholme et al., 2010; Barros et al., 2015; Cui et al., 2022; Ghosh et al., 2022; Yin et al., 2022). Our comprehensive analysis of the soybean CCR gene family provides novel insights into the evolutionary expansion and functional diversification of this enzyme family in legumes, revealing significant differences from previously characterized plant species.

4.1 Evolutionary expansion and phylogenetic relationships

Our identification of 15 *GmCCR* genes represents a notable expansion compared to previous studies in other plant species. This number exceeds the 11 members reported in *Arabidopsis thaliana* (Costa et al., 2003) and is comparable to the 13 members identified in flax (*Linum usitatissimum*) (Huis et al., 2012) and the 10 members in *Dalbergia odorifera* (Wang et al., 2022). The phylogenetic analysis confirmed the established four-subfamily classification (Ia, Ib, Ic, and II), with seven *DoCCRs* grouped with functionally characterized CCRs of dicotyledons involved in developmental lignification, demonstrating the evolutionary conservation of CCR gene organization across plant lineages. Importantly, our phylogenetic reconstruction revealed that soybean CCR genes are distributed across all four subfamilies, with a particularly notable expansion in Subfamily II (containing 12 members). This contrasts with the distribution patterns observed in *Arabidopsis*, where Subfamily II contains only 9 members (Costa et al., 2003), and in rice, where *OsCCR20* and 19 were grouped with known plant CCRs but showed more even distribution across subfamilies (Kawasaki et al., 2006). The preferential expansion of Subfamily II in soybean suggests legume-specific evolutionary pressures that favored the retention and diversification of these potentially multifunctional CCR-like proteins. Previous studies have established that *SbCCR1* was closer to other CCR1 proteins involved in lignin biosynthesis in plant developmental processes in sorghum (Sattler et al., 2010). Similar patterns have been observed across multiple species. Our soybean sequences fit well within this established evolutionary framework, with four *GmCCR* genes (*GmCCR2*, *GmCCR4*, *GmCCR10*, and *GmCCR12*) clustering with functionally characterized *AtCCR1* and *AtCCR2*, providing strong evidence for their roles in lignin biosynthesis.

The diversity in predicted subcellular localizations represents a previously unreported feature in CCR gene families and may reflect the complex cellular requirements for lignin biosynthesis in legumes. The dual targeting potential of *GmCCR7, GmCCR10*, and *GmCCR15* (showing both chloroplast and cytoplasmic/Golgi predictions) is particularly intriguing and may represent a legume-specific adaptation. This dual localization could serve multiple functional roles: chloroplast localization may support lignin precursor synthesis in photosynthetic tissues where carbon skeletons are readily available, Golgi apparatus localization aligns with the traditional role in lignin monomer processing and cell wall transport, and cytoplasmic localization may enable rapid stress responses through direct interaction with stress signaling pathways. Furthermore, the specialized cellular environments in legume root nodules, where symbiotic nitrogen fixation creates unique metabolic demands, may require flexible CCR localization to support both structural (infection thread formation) and defense-related lignification. The dual targeting may allow these proteins to respond dynamically to cellular conditions, shifting between compartments based on metabolic needs or stress signals. This compartmental flexibility could provide evolutionary advantages in the complex cellular environment's characteristic of legume-rhizobia symbiosis.

4.2 Gene duplication patterns and synteny conservation

The synteny analysis revealed 12 collinear gene pairs within soybean, indicating that segmental duplication events significantly contributed to CCR family expansion. This pattern is consistent with the well-documented paleopolyploidy events in soybean evolutionary history (Schmutz et al., 2010), where approximately 75% of genes exist as duplicates resulting from a whole-genome duplication event ~13 million years ago, followed by extensive gene loss and subfunctionalization (Shoemaker et al., 2006). The distribution of duplicated gene pairs across subfamilies provides insights into evolutionary constraints and functional importance. The equal distribution of duplicated pairs between Subfamilies Ia and II (6 pairs each) suggests that both true CCRs and CCR-like proteins experienced similar evolutionary pressures following duplication events, potentially indicating comparable functional importance in soybean biology. Comparative analysis with other legumes reveals interesting patterns identical to those observed in *Medicago truncatula* and other members of the Fabaceae family (Young et al., 2011; Tang et al., 2014). Cross-species synteny analysis between dicotyledonous plants has identified orthologous relationships for stress-responsive genes, and our analysis extends these findings to show conservation of CCR gene organization across plant families. The maintenance of syntenic relationships suggests that CCR genes occupy critical regulatory positions in plant genomes, with their chromosomal context potentially important for proper expression regulation.

4.3 Functional diversification and subfamily-specific roles

The motif analysis revealed interesting patterns of conservation and divergence among soybean CCR genes. While all 15 members contain the five core motifs essential for CCR enzymatic activity, the differential presence of motifs 7 and 9 in Subfamily II members

suggests functional diversification. This pattern is consistent with recent studies in flax, where LuCCR13/20 were found to align closely with functional CCRs involved in lignin biosynthesis in dicotyledonous plants and share NADP-specificity, NAD(P)-B, and CCR signature motifs with known functional CCRs (Huis et al., 2012; Chao et al., 2017; Song et al., 2025). Similar functional diversification has been reported in *Liriodendron chinense*, where LcCCR13 revealed potential roles extending beyond traditional lignin biosynthesis (Li et al., 2023). The subcellular localization predictions revealed an interesting distribution pattern not previously reported in other species. While most CCR genes encode proteins targeted to the Golgi apparatus (consistent with their role in lignin precursor synthesis), the presence of cytoplasmic and dual-localized proteins suggests additional cellular functions. This diversity in subcellular targeting is particularly notable in legumes, where specialized cell types and symbiotic relationships may require CCR activity in multiple cellular compartments (Radwan et al., 2011; Chao et al., 2019).

4.4 Tissue-specific expression and developmental regulation

Our transcriptome analysis revealed tissue-specific expression patterns that both confirm and extend previous findings. The high expression of most *GmCCR* genes in roots and stems aligns with their expected role in lignification and structural support, consistent with findings in *Arabidopsis* and rice (Tamasloukht et al., 2011; Seo-Won, 2019; Yin et al., 2021). However, the notable expression in developing seeds represents a potentially legume-specific feature, as seed lignification is particularly important in legume species for seed coat development and protection. The differential expression patterns among family members suggest functional specialization that extends beyond simple redundancy. *GmCCR12*'s predominant root expression and *GmCCR9*'s stem-specific expression indicate that gene duplication events were followed by subfunctionalization, allowing for tissue-specific optimization of CCR activity (Birchler and Yang, 2022). This pattern is consistent with the neofunctionalization model proposed for duplicated genes in plant families (Porth et al., 2011).

4.5 Stress responsiveness and regulatory networks

Our stress expression analysis revealed that soybean CCR genes exhibit more complex stress responses than previously characterized in other species. The identification of five genes (*GmCCR1*, *GmCCR4*, *GmCCR7*, *GmCCR8*, and *GmCCR15*) that are consistently upregulated across all stress treatments represents a novel finding in CCR biology. This broad stress responsiveness contrasts with the more specific responses reported in *Arabidopsis* and rice, suggesting that legumes may have evolved enhanced stress tolerance mechanisms involving CCR-mediated pathways. The temporal expression patterns, particularly the rapid upregulation of *GmCCR8* at 3 hours post-treatment, indicate that CCR genes function as early-response elements in stress signaling cascades. This rapid response is consistent with the role of phenylpropanoid metabolism in immediate stress defense, including the production of protective compounds and cell wall modifications. The comprehensive promoter analysis revealed a complex regulatory landscape with multiple hormone-responsive elements. The presence of ABA-responsive elements in most promoters aligns with the observed upregulation under osmotic stress conditions. It is consistent with recent findings in banana, where lignin biosynthesis genes showed essential roles in fruit ripening and stress response. The diversity of cis-regulatory elements suggests that CCR genes are integrated into multiple regulatory networks, allowing for coordinated responses to diverse environmental signals. The abundance of specific cis-regulatory elements in highly stress-responsive genes provides mechanistic insights into their regulation. *GmCCR4*, which showed the most dramatic early stress responses, contains 6 ABRE (ABA-responsive elements), 4 MeJA-responsive elements, and 3 drought-responsive elements (DRE) in its promoter region. This high density of stress-responsive elements correlates directly with its broad stress responsiveness and early activation kinetics (peak expression at 3h post-treatment).Similarly, *GmCCR8*, which exhibited the highest peak expression levels (40-fold increase), possesses 5 ABRE elements and multiple TC-rich repeats associated with defense responses, explaining its sustained upregulation across all stress treatments. The presence of both ABA-dependent (ABRE) and ABA-independent (DRE) elements in these promoters suggests dual regulatory pathways that enable rapid initial responses through ABA-independent mechanisms, followed by sustained expression through ABA-dependent signaling. In contrast, the consistently downregulated genes *GmCCR2* and *GmCCR11* lack multiple stress-responsive elements but contain numerous auxin-responsive elements (AuxRE), suggesting their primary roles in developmental processes that are suppressed during stress to redirect metabolic resources toward stress tolerance mechanisms. The consistent downregulation of *GmCCR2* and *GmCCR11* across all stress treatments suggests important regulatory roles that extend beyond simple loss of function. Several mechanisms may explain this negative regulation pattern: Metabolic resource reallocation: Downregulation of these genes may redirect carbon flux and cellular resources away from normal developmental lignification toward stress-specific defensive compounds and osmolytes. Cell wall remodeling specificity: These genes may typically produce lignin precursors for specific cell wall layers or tissue types that become counterproductive under stress conditions, requiring their suppression to allow stress-adaptive cell wall modifications. Temporal regulation hierarchy: GmCCR2 and GmCCR11 may function primarily during non-stress conditions to maintain fundamental structural integrity, while stress conditions activate alternative CCR genes (*GmCCR1, GmCCR4*, *GmCCR8*) optimized for rapid defensive responses. Substrate competition prevention:

Active downregulation may prevent these enzymes from competing with stress-responsive CCRs for shared substrates, ensuring efficient channeling of phenylpropanoid precursors toward stress-protective compounds. This regulatory strategy resembles the 'metabolic switching' observed in other stress-responsive pathways, where normal housekeeping enzymes are suppressed to favor stress-specific isoforms with different kinetic properties or substrate specificities optimized for stress conditions".

4.6 Implications for legume biology and crop improvement

The expansion and diversification of the CCR gene family in soybean have essential implications for legume biology and agriculture. The increased stress responsiveness of multiple family members offers molecular targets for creating stress-tolerant soybean varieties, which is especially crucial given the rising challenges of climate change and soil salinity in farming. The tissue-specific expression patterns indicate that different CCR genes could be targeted for specific improvements: root-expressed genes for better stress tolerance and nutrient absorption, stem-expressed genes for stronger lodging resistance and water transport, and seed-expressed genes for improved seed quality and storage protein content.

4.7 Limitations of the study

While our comprehensive genomic and transcriptomic analysis provides valuable insights into CCR gene family evolution and stress responses, several limitations should be acknowledged: (1) Functional validation gap: Our study relies primarily on expression analysis without direct experimental validation of protein function or stress tolerance improvement through genetic modification. Future studies should prioritize functional validation through overexpression, knockdown, and genome editing approaches coupled with physiological assessments of lignin content and stress resilience. (2) Single-tissue analysis limitation: Our stress expression analysis focused on root tissues, while CCR genes may have tissue-specific stress responses that could provide additional insights into their functional specialization. Multi-tissue stress analysis would provide a more comprehensive understanding of CCR family roles in whole-plant stress responses. (3) Subfamily II functional characterization: The notable expansion of Subfamily II in soybean (12 members) represents an evolutionary innovation that requires deeper functional characterization. While our structural and expression analyses provide initial insights, the specific functions and potential neo-functionalization of these expanded members remain to be experimentally determined. (4) Mechanistic details: While we identify stress-responsive cis-elements and correlate them with expression patterns, the specific transcription factors and signaling pathways mediating these responses require further investigation through protein-DNA interaction studies and regulatory network analysis.

5 Conclusions

This study identifies 15 members across 12 chromosomes with varied origins and functions. Phylogenetic analysis revealed four main subfamilies linked to CCR genes in other plants, while synteny showed segmental duplications aided in family expansion. Diverse cis-regulatory elements in *GmCCR* promoters and tissue-specific, stress-responsive expression patterns indicate a complex regulatory network. Five genes (*GmCCR1, GmCCR4, GmCCR7, GmCCR8*, and *GmCCR15*) increased expression under salt, alkaline, and osmotic stresses, suggesting roles in abiotic stress tolerance. Root-specific and stress-responsive expression links lignin biosynthesis and phenylpropanoid metabolism to adaptation in harsh soils, relevant for China's saline-alkali soils affecting soybean growth amid rising demand. Our findings offer a molecular framework for soybean's cell wall response to stress and identify gene targets for crop improvement. Stress-responsive *GmCCR* genes are promising for marker-assisted selection or genetic engineering to develop salt-alkali-tolerant soybeans. This research advances understanding of CCR evolution and function in legumes and provides tools for sustainable farming. As climate change and soil degradation threaten agriculture, these insights and resources are vital for creating resilient crops capable of maintaining productivity in stressful environments.

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

XL: Writing – review & editing, Project administration, Investigation, Conceptualization, Software, Resources, Writing – original draft. YL: Methodology, Data curation, Writing – review & editing, Writing – original draft, Software, Visualization, Funding acquisition. SiL: Formal Analysis, Writing – review & editing, Resources, Supervision, Methodology, Conceptualization, Writing – original draft. MS: Writing – original draft, Software, Conceptualization, Investigation, Formal Analysis, Writing – review & editing, Project administration. QC: Methodology, Conceptualization, Validation, Funding acquisition, Writing – original draft, Writing – review & editing, Investigation. YS: Supervision, Conceptualization, Writing – review & editing, Validation, Writing – original draft, Software, Visualization. ShL: Supervision, Writing – review & editing, Validation, Methodology, Funding acquisition, Formal Analysis, Resources, Writing – original draft, Data curation. YY: Software, Writing – review & editing, Funding acquisition, Writing – original

draft, Conceptualization, Formal Analysis, Data curation. TY: Project administration, Conceptualization, Writing – review & editing, Investigation, Methodology, Formal Analysis, Writing – original draft, Data curation. JZ: Writing – original draft, Resources, Investigation, Funding acquisition, Visualization, Writing – review & editing, Validation, Conceptualization, Supervision.

Funding

The author(s) declare financial support was received for the research and/or publication of this article. This study was financially supported by Heilongjiang Provincial Touyan Project, National Frigid Zone Crop and Soybean Germplasm Resources Mid-term Bank, Innovation Project of Heilongjiang Academy of Agricultural Sciences (CX23ZD05,CX23JQ04), Biological Breeding-National Science and Technology Major Project (2023ZD04027), National Key R&D Program of China (2021YFD1201000).

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.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2025.1657111/full#supplementary-material

References

Aoyagi, L. N., Lopes-Caitar, V. S., De Carvalho, M. C., Darben, L. M., Polizel-Podanosqui, A., Kuwahara, M. K., et al. (2014). Genomic and transcriptomic characterization of the transcription factor family R2R3-MYB in soybean and its involvement in the resistance responses to Phakopsora pachyrhizi. *Plant Sci.* 229, 32–42. doi: 10.1016/j.plantsci.2014.08.005

Barros, J., Serk, H., Granlund, I., and Pesquet, E. (2015). The cell biology of lignification in higher plants. *Ann. Bot.* 115, 1053–1074. doi: 10.1093/aob/mcv046

Birchler, J. A., and Yang, H. (2022). The multiple fates of gene duplications: deletion, hypofunctionalization, subfunctionalization, neofunctionalization, dosage balance constraints, and neutral variation. *Plant Cell* 34, 2466–2474. doi: 10.1093/plcell/koac076

Cannon, S. B., and Shoemaker, R. C. (2012). Evolutionary and comparative analyses of the soybean genome. *Breed. Sci.* 61, 437–444. doi: 10.1270/jsbbs.61.437

Chao, N., Jiang, W.-T., Wang, X.-C., Jiang, X.-N., and Gai, Y. (2019). Novel motif is capable of determining CCR and CCR-like proteins based on the divergence of CCRs in plants. *Tree Physiol.* 39, 2019–2026. doi: 10.1093/treephys/tpz098

Chao, N., Li, N., Qi, Q., Li, S., Lv, T., Jiang, X.-N., et al. (2017). Characterization of the cinnamoyl-CoA reductase (CCR) gene family in Populus tomentosa reveals the enzymatic active sites and evolution of CCR. *Planta* 245, 61–75. doi: 10.1007/s00425-016-2591-6

Chun, H. J., Baek, D., Cho, H. M., Lee, S. H., Jin, B. J., Yun, D.-J., et al. (2019). Lignin biosynthesis genes play critical roles in the adaptation of Arabidopsis plants to high-salt stress. *Plant Signaling Behav.* 14, 1625697. doi: 10.1080/15592324.2019.1625697

Costa, M. A., Collins, R. E., Anterola, A. M., Cochrane, F. C., Davin, L. B., and Lewis, N. G. (2003). An in silico assessment of gene function and organization of the phenylpropanoid pathway metabolic networks in Arabidopsis thaliana and limitations thereof. *Phytochemistry* 64, 1097–1112. doi: 10.1016/S0031-9422(03)00517-X

Cui, W., Zhuang, Z., Jiang, P., Pan, J., Zhao, G., Xu, S., et al. (2022). Characterization, expression profiling, and biochemical analyses of the cinnamoyl-CoA reductase gene family for lignin synthesis in alfalfa plants. *Int. J. Mol. Sci.* 23, 7762. doi: 10.3390/ijms23147762

Do, T. D., Vuong, T. D., Dunn, D., Clubb, M., Valliyodan, B., Patil, G., et al. (2019). Identification of new loci for salt tolerance in soybean by high-resolution genome-wide association mapping. *BMC Genomics* 20, 1–16. doi: 10.1186/s12864-019-5662-9

Ghosh, S., Zhang, S., Azam, M., Gebregziabher, B. S., Abdelghany, A. M., Shaibu, A. S., et al. (2022). Natural variation of seed tocopherol composition in diverse world soybean accessions from maturity group 0 to VI grown in China. *Plants* 11, 206. doi: 10.3390/plants11020206

Huang, S., Kang, X., Fu, R., Zheng, L., Li, P., Tang, F., et al. (2024). Simultaneous down-regulation of dominant cinnamoyl coA reductase and cinnamyl alcohol dehydrogenase dramatically altered lignin content in mulberry. *Plants* 13, 3512. doi: 10.3390/plants13243512

Huis, R., Morreel, K., Fliniaux, O., Lucau-Danila, A., Fénart, S., Grec, S., et al. (2012). Natural hypolignification is associated with extensive oligolignol accumulation in flax stems. *Plant Physiol.* 158, 1893–1915. doi: 10.1104/pp.111.192328

Islam, I., Adam, Z., and Islam, S. (2019). Soybean (Glycine Max): alternative sources of human nutrition and bioenergy for the 21st century. *Am. J. Food Sci. Technol.* 7, 1–6. doi: 10.12691/ajfst-7-1-1

Kawasaki, T., Koita, H., Nakatsubo, T., Hasegawa, K., Wakabayashi, K., Takahashi, H., et al. (2006). Cinnamoyl-CoA reductase, a key enzyme in lignin biosynthesis, is an effector of small GTPase Rac in defense signaling in rice. *Proc. Natl. Acad. Sci.* 103, 230–235. doi: 10.1073/pnas.0509875103

Lamlom, S. F., Zhang, Y., Su, B., Wu, H., Zhang, X., Fu, J., et al. (2020). Map-based cloning of a novel QTL qBN-1 influencing branch number in soybean [Glycine max (L.) Merr. *Crop J.* 8, 793–801. doi: 10.1016/j.cj.2020.03.006

Li, W., Hao, Z., Yang, L., Xia, H., Tu, Z., Cui, Z., et al. (2023). Genome-wide identification and characterization of LcCCR13 reveals its potential role in lignin biosynthesis in Liriodendron chinense. *Front. Plant Sci.* 13, 1110639. doi: 10.3389/fpls.2022.1110639

Li, J. W., Zhou, P., Hu, Z. H., Teng, R. M., Wang, Y. X., Li, T., et al. (2024). CsPAT1, a GRAS transcription factor, promotes lignin accumulation by antagonistic interacting with CsWRKY13 in tea plants. *Plant J.* 118, 1312–1326. doi: 10.1111/tpj.16670

Liu, C.-J. (2012). Deciphering the enigma of lignification: precursor transport, oxidation, and the topochemistry of lignin assembly. *Mol. Plant* 5, 304–317. doi: 10.1093/mp/ssr121

Liu, D., Wu, J., Lin, L., Li, P., Li, S., Wang, Y., et al. (2021). Overexpression of Cinnamoyl-CoA Reductase 2 in Brassica napus increases resistance to Sclerotinia sclerotiorum by affecting lignin biosynthesis. *Front. Plant Sci.* 12, 732733. doi: 10.3389/fpls.2021.732733

Ma, Q.-H. (2024). Lignin biosynthesis and its diversified roles in disease resistance. *Genes* 15, 295. doi: 10.3390/genes15030295

Modgil, R., Tanwar, B., Goyal, A., and Kumar, V. (2021). Soybean (glycine max). In Tanwar, B., and Goyal, A. (eds) *Oilseeds: health attributes and food applications* (Singapore: Springer). doi: 10.1007/978-981-15-4194-0_1

Muro-Villanueva, F., Kim, H., Ralph, J., and Chapple, C. (2022). H-lignin can be deposited independently of CINNAMYL ALCOHOL DEHYDROGENASE C and D in Arabidopsis. *Plant Physiol.* 189, 2015–2028. doi: 10.1093/plphys/kiac210

Neves, G., Marchiosi, R., Ferrarese, M., Siqueira-Soares, R., and Ferrarese-Filho, O. (2010). Root growth inhibition and lignification induced by salt stress in soybean. *J. Agron. Crop Sci.* 196, , 467–, 473. doi: 10.1111/j.1439-037X.2010.00432.x

Nizam, A., Thattantavide, A., and Kumar, A. (2024). Gene expression pattern, lignin deposition and root cell wall modification of developing mangrove propagules under salinity stress. *J. Plant Growth Regul.* 43, 3088–3104. doi: 10.1007/s00344-023-11021-z

Pb, K. K., Singam, P., and Suravajhala, P. (2023). Modulation of lignin and its implications in salt, drought and temperature stress tolerance. *Curr. Chem. Biol.* 17, 2–12. doi: 10.2174/2212796816666220820110616

Porth, I., Hamberger, B., White, R., and Ritland, K. (2011). Defense mechanisms against herbivory in Picea: sequence evolution and expression regulation of gene family members in the phenylpropanoid pathway. *BMC Genomics* 12, 1–26. doi: 10.1186/1471-2164-12-608

Qiu, L.-J., Xing, L.-L., Guo, Y., Wang, J., Jackson, S. A., and Chang, R.-Z. (2013). A platform for soybean molecular breeding: the utilization of core collections for food security. *Plant Mol. Biol.* 83, 41–50. doi: 10.1007/s11103-013-0076-6

Radwan, O., Liu, Y., and Clough, S. J. (2011). Transcriptional analysis of soybean root response to Fusarium virguliforme, the causal agent of sudden death syndrome. *Mol. Plant-Microbe Interact.* 24, 958–972. doi: 10.1094/MPMI-11-10-0271

Ren, H., Zhang, B., Zhang, F., Liu, X., Wang, X., Zhang, C., et al. (2024). Integration of physiological and transcriptomic approaches in investigating salt-alkali stress resilience in soybean. *Plant Stress* 11, 100375. doi: 10.1016/j.stress.2024.100375

Sattler, S. E., Funnell-Harris, D. L., and Pedersen, J. F. (2010). Brown midrib mutations and their importance to the utilization of maize, sorghum, and pearl millet lignocellulosic tissues. *Plant Sci.* 178, 229–238. doi: 10.1016/j.plantsci.2010.01.001

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

Seo-Won, C. (2019). *Characterization of the osCCR10, a lignin biosynthesis gene involved in rice drought tolerance mechanism* (Seoul National University Graduate School). Available online at: https://s-space.snu.ac.kr/handle/10371/161109

Severin, A. J., Woody, J. L., Bolon, Y.-T., Joseph, B., Diers, B. W., Farmer, A. D., et al. (2010). RNA-Seq Atlas of Glycine max: a guide to the soybean transcriptome. *BMC Plant Biol.* 10, 1–16. doi: 10.1186/1471-2229-10-160

Shahzad, A., Ullah, S., Dar, A. A., Sardar, M. F., Mehmood, T., Tufail, M. A., et al. (2021). Nexus on climate change: Agriculture and possible solution to cope future climate change stresses. *Environ. Sci. pollut. Res.* 28, 14211–14232. doi: 10.1007/s11356-021-12649-8

Shoemaker, R. C., Schlueter, J., and Doyle, J. J.. (2006). Paleopolyploidy and gene duplication in soybean and other legumes. *Current opinion in plant biology* 9 (2), 104–109.

So, H.-A., Chung, E., Cho, C.-W., Kim, K.-Y., and Lee, J.-H. (2010). Molecular cloning and characterization of soybean cinnamoyl CoA reductase induced by abiotic stresses. *Plant Pathol. J.* 26, 380–385. doi: 10.5423/PPJ.2010.26.4.380

Song, X., Liu, D., Yao, Y., Tang, L., Cheng, L., Yang, L., et al. (2025). Genome-wide identification and expression pattern analysis of the cinnamoyl-CoA reductase gene family in flax (Linum usitatissimum L.). *BMC Genomics* 26, 315. doi: 10.1186/s12864-025-11481-5

Tamasloukht, B., Wong Quai Lam, M. S.-J., Martinez, Y., Tozo, K., Barbier, O., Jourda, C., et al. (2011). Characterization of a cinnamoyl-CoA reductase 1 (CCR1) mutant in maize: effects on lignification, fibre development, and global gene expression. *J. Exp. Bot.* 62, 3837–3848. doi: 10.1093/jxb/err077

Tang, H., Krishnakumar, V., Bidwell, S., Rosen, B., Chan, A., Zhou, S., et al. (2014). An improved genome release (version Mt4. 0) for the model legume Medicago truncatula. *BMC Genomics* 151–14. doi: 10.1186/1471-2164-15-312

Vanholme, R., Demedts, B., Morreel, K., Ralph, J., and Boerjan, W. (2010). Lignin biosynthesis and structure. *Plant Physiol.* 153, 895–905. doi: 10.1104/pp.110.155119

Wang, Y., Xu, J., Zhao, W., Li, J., and Chen, J. (2022). Genome-wide identification, characterization, and genetic diversity of CCR gene family in Dalbergia odorifera. *Front. Plant Sci.* 13, 1064262. doi: 10.3389/fpls.2022.1064262

Yin, N., Li, B., Liu, X., Liang, Y., Lian, J., Xue, Y., et al. (2021). Cinnamoyl-coA reductase 1 (CCR1) and CCR2 function divergently in tissue lignification, flux control and cross-talk with glucosinolate pathway in brassica napus. *bioRxiv* 2003, 2001.433400. doi: 10.1101/2021.03.01.433400

Yin, N., Li, B., Liu, X., Liang, Y., Lian, J., Xue, Y., et al. (2022). Two types of cinnamoyl-CoA reductase function divergently in accumulation of lignins, flavonoids and glucosinolates and enhance lodging resistance in Brassica napus. *Crop J.* 10, 647–660. doi: 10.1016/j.cj.2021.10.002

Young, N. D., Debellé, F., Oldroyd, G. E., Geurts, R., Cannon, S. B., Udvardi, M. K., et al. (2011). The Medicago genome provides insight into the evolution of rhizobial symbioses. *nature* 480, 520–524. doi: 10.1038/nature10625

Zheng, J., Sun, L., Wang, D., He, L., Du, W., Guo, S., et al. (2023). Roles of a CCR4–NOT complex component GmNOT4–1 in regulating soybean nodulation. *Front. Plant Sci.* 14, 1172354. doi: 10.3389/fpls.2023.1172354

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 28 November 2025
DOI 10.3389/fpls.2025.1684712

OPEN ACCESS

EDITED BY
Gurjeet Singh,
Texas A&M University, United States

REVIEWED BY
Prabina Kumar Meher,
Indian Agricultural Statistics Research Institute, Indian Council of Agricultural Research (ICAR), India
Dinesh Kumar Saini,
Texas Tech University, United States

*CORRESPONDENCE
Niranjan Baisakh
nbaisakh@agcenter.lsu.edu
Valarmathi Ramanathan
valarmathi.r@icar.org.in

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

RECEIVED 13 August 2025
ACCEPTED 14 October 2025
PUBLISHED 28 November 2025

CITATION
Ramanathan V, Pradhan AK, Gandham P, Chinnaswamy A, Mahadeva Swamy HK, Mohanraj K, R R and Baisakh N (2025) Identification of new genetic resources for drought tolerance-related traits from the world *Erianthus* germplasm collection. *Front. Plant Sci.* 16:1684712. doi: 10.3389/fpls.2025.1684712

Identification of new genetic resources for drought tolerance-related traits from the world *Erianthus* germplasm collection

Valarmathi Ramanathan[1*†], Anjan Kumar Pradhan[2†], Prasad Gandham[2], Appunu Chinnaswamy[1], H. K. Mahadeva Swamy[1], K. Mohanraj[1], Rasitha R[1] and Niranjan Baisakh[2*]

[1]Division of Crop Improvement, Indian Council of Agricultural Research (ICAR)-Sugarcane Breeding Institute, Coimbatore, India, [2]School of Plant, Environmental and Soil Sciences, Louisiana State University Agricultural Center, Baton Rouge, LA, United States

Drought is one of the most complex abiotic stress factors and significantly affects sugarcane production worldwide. With an objective to identify drought-tolerant wild sugarcane germplasm for sugarcane improvement, a set of 223 diverse *Erianthus* clones collected from seven different countries were assessed for drought tolerance-related traits at the formative stage. A drought association panel of 91 *Erianthus* clones with differential drought response was developed for genomic studies to identify marker–trait associations. Drought-tolerant *Erianthus* clones maintained high chlorophyll fluorescence (F_v/F_m) and relative water content and chlorophyll content, and were able to regulate photosynthetic rate as well as canopy cooling to sustain growth and biomass accumulation. They also recorded high stress tolerance indices and a low susceptibility index. Moderate to high heritability estimates of the traits suggested that the traits were under genetic control. A genome-wide association study conducted using 1,044 genotyping by sequencing-derived high-quality single-nucleotide polymorphism markers on the drought panel identified 43 quantitative trait nucleotides (QTNs) associated with various physiological traits that explained phenotypic variance ranging from 0.26% to 59.35%. Candidate genes [Target of Rapamycin 2 (TOR2), Trans-Membrane Kinase 1 (TMK1), potassium transporter auxin response factor, FAD-binding PCMH-type domain-containing protein, and Nitrate Transporter 1/Peptide Transporter Family (NRT1/PTR)] involved in drought perception and signaling were successfully identified in QTN regions. The present study identified drought-tolerant *Erianthus* clones and the QTNs associated with drought tolerance traits for further validation and subsequent utilization in the sugarcane breeding program.

KEYWORDS

drought, *Erianthus*, heritability, marker–trait association, SNP effect

Introduction

The diverse wild germplasm collection is an important source of desirable alleles contributing to improved yield and resistance to various biotic and abiotic stresses (Fukuhara et al., 2013). Sugarcane (*Saccharum* spp. hybrids) belongs to the genus *Saccharum*, which, along with potential interbreeding genera (*Erianthus, Miscanthus, Narenga*, and *Sclerostachya*), forms the "*Saccharum* complex" and serves as the major wild germplasm for sugarcane (Mukherjee, 1957).

Erianthus is the most closely related genus to the *Saccharum* primary gene pool (Mukherjee, 1958; Nair and Mary, 2006). Among seven closely related species of *Erianthus* (*E. arundinaceus*, *E. bengalense*, *E. elephantinus*, *E. hosti*, *E. ravennae*, *E. procerus*, and *E. kanashiroi*), *E arundinaceus* is the most widely distributed species reported from India, Burma, China, Indonesia, Malaysia, New Guinea, Philippines, and Thailand (Mukherjee, 1957; Dao et al., 2013; Tsuruta et al., 2022). All *Erianthus* species are known for their exceptional adaptability to abiotic and biotic stress, high biomass, high tillering ability, and better ratooning ability, which have been attributed to its extensive root system (Nair et al., 2017; Wang et al., 2019; Valarmathi et al., 2020, 2023; Meena et al., 2024; Takaragawa and Wakayama, 2024; Wang et al., 2025).

Intergeneric hybridization of sugarcane with other related genera within the *Saccharum* complex was attempted to broaden the genetic base of modern sugarcane cultivars and introgress more productive and better stress adaptation traits (Nair et al., 2017; Piperidis et al., 2000; Terajima et al., 2022, 2023). Early hybridization attempts between *Saccharum* spp. and *Erianthus* as a potential breeding material for improving sugarcane resulted in limited success due to cross-incompatibility, inherent difficulty in identifying true hybrids from self-progeny, chromosome elimination, and unreduced gamete formation (D'Hont et al., 1995; Piperidis and D'hont, 2001; Piperidis et al., 2010; Jackson and Henry, 2011). However, consequent hybridization efforts led to a few successful intergeneric hybrids between sugarcane and *Erianthus* (Deng et al., 2002; Cai et al., 2005; Piperidis et al., 2010; Chang et al., 2012; Fukuhara et al., 2013; Nair et al., 2017; Pachakkil et al., 2019; Meena et al., 2024). Hybrids developed from the cross between *Erianthus procerus* "IND 90-776" and *Saccharum officinarum* "PIO 96-435" showed higher resistance to red rot with better tolerance to drought (Nair et al., 2017). Progeny of sugarcane and *E. arundinaceus* possessed improved root characteristics and enhanced drought tolerance traits (Fukuhara et al., 2013), resistance to low temperature, and red rot resistance (Ram et al., 2001). *Erianthus*, as a high biomass plant, is also utilized in developing high-fiber energy canes as a bioenergy crop (Deng et al., 2002; Chang et al., 2012; Meena et al., 2024). Several hybridization attempts using *Saccharum* and *E. arundinaceus* as a male parent successfully produced hybrids with desirable traits transferred from *Erianthus* (Sreenivasan et al., 1987; Ram et al., 2001).

Drought is a major abiotic stress factor limiting sugarcane production worldwide (Wang et al., 2003; Dhansu et al., 2021). Most commercial sugarcane varieties, with their shallow root system, are susceptible to moisture stress as well as other belowground extreme conditions (Pierre et al., 2019), which results in reduction in shoot growth, cane yield, and sucrose content (Hoang et al., 2019). World germplasm collections of *E. arundinaceus* clones collected from all over India and other countries such as Burma, Fiji, Indonesia, New Guinea, Pakistan, and Philippines maintained at the Indian Council of Agricultural Research–Sugarcane Breeding Institute, Coimbatore represent a diversified *Saccharum* germplasm resource (Cai et al., 2012; Scortecci et al., 2012; Valarmathi et al., 2021). *Erianthus* species clones with a deeper and prolific root system architecture have been shown to impart drought tolerance by leveraging more underground resources (Narayanan et al., 2014; Valarmathi et al., 2020, 2023). However, sequence variants controlling drought tolerance among *Erianthus* genetic resources have not been identified for their use in targeted sugarcane breeding programs.

Despite the high biomass, good Brix value, ratooning ability, and wide environmental adaptability, genetic studies in *Erianthus* are very limited (Cai et al., 2005; Tsuruta et al., 2012, 2022). On the other hand, genotyping by sequencing (GBS)-derived single-nucleotide polymorphisms (SNPs) have been used for the identification of markers linked to genes associated with various agronomic traits of interest in sugarcane via QTL mapping using biparental or self-population and/or genome-wide association mapping using a diversity panel (Balsalobre et al., 2017; Yang et al., 2018, 2019, 2020; Gutierrez et al., 2018; Fickett et al., 2019; Wirojsirasak et al., 2023; Cortes et al., 2024). The recently available reference genome of the wild progenitor *S. spontaneum*, sugarcane hybrids, and sequence variations can be better identified at high resolution in sugarcane (Garsmeur et al., 2018; Zhang et al., 2018; Bao et al., 2024; Healey et al., 2024). Some *Erianthus* species such as *E. fulvus* and *E. rufipilus* possess the lowest chromosome number (2n=2x=20) within the *Saccharum* complex (Amalraj and Balasundaram, 2006) and their genome is now available to facilitate genetic studies (Kui et al., 2023; Wang et al., 2023).

In this first ever genome-wide association study, 223 accessions from the world germplasm collection of *Erianthus* maintained at the Sugarcane Breeding Institute, Coimbatore, India were phenotyped for drought tolerance-related traits from 2017 to 2020. Based on the 3 years of field phenotyping data, a drought panel of 91 clones was constructed for a genome-wide association analysis, which identified 43 significant quantitative trait nucleotides (QTNs) and linked genes associated with 12 traits (control and drought stress).

Materials and methods

Plant material

A total of 223 germplasm accessions belonging to five different species of the genus *Erianthus* [*E. arundinaceus* (208) clones, *E. bengalense* (3), *E. procerus* (5), *E. ravennae* (3), and *E. elegans* (4)] maintained at the ICAR, Sugarcane Breeding Institute, Coimbatore, India were used in the study. The germplasm included collections from different geographical regions from countries such as Burma,

Fiji, India, Indonesia, New Guinea, Pakistan, and Philippines (Valarmathi et al., 2021).

Experimental site and climatic conditions

The experiment was conducted at the farm field of ICAR Sugarcane Breeding Institute, Coimbatore, India (77°E longitude and 11°N latitude; elevation, 427 m) (1,500–1,800 μmol m^{-2} s^{-1} light intensity, photoperiod of 16 h light and 8 h dark, temperature of 30°C ± 2°C with ~75% relative humidity). The soil at the experimental site was sandy clay loam, taxonomically classified as typic haplustalf with a pH of 7.7. The field site experiences a typical tropical wet and dry climate, experiencing dry season from March to June, intermittent drying from July to September, and wet season lasting from October to December due to the northeast monsoon. Planting was done in December, and the standard agronomic cultural method was followed, which included applications of N, P, and K at 280, 62.5, and 120 kg ha^{-1}, respectively. Phosphorus (62.5 kg ha^{-1}) was applied as basal dressing before planting, while nitrogen and potassium were applied at 45 and 90 days after planting in two equal splits.

Experimental design and drought stress

The field evaluation of drought tolerance of the 223 *Erianthus* accessions was laid out in an augmented block design (ABD) trial for 2 years (2017–2018) along with drought-tolerant (SES 288 and SES 293) and drought-susceptible (IS 76–215 and IS 76 218) checks. From the ABD trial, 91 *Erianthus* clones were selected and evaluated in a randomized block design (RBD) trial for 3 years (2019–2021) in three replications along with the drought-tolerant and -susceptible checks. For both ABD and RBD trials, the clones were planted in 6-m-long rows with 3 m spacing between the rows and 0.5 m spacing between clumps. Two-budded setts of each clone were planted at 12 setts for every 6-m-long row with 3 m spacing between the rows and 0.5 m spacing between clumps. Each plot had 20 rows, 60 m long with 2 m spacing between adjacent plots and 5 m spacing between control and drought blocks.

Drought stress was imposed by withholding irrigation at the formative phase of the crop (60 days after planting). The progress of drought stress was monitored by measuring the soil moisture. The drought trial was left completely under rainfed conditions throughout the crop cycle, and the biomass data were recorded at harvest. The soil moisture content of the experimental field was measured using a pressure plate apparatus (Chakraborty et al., 2010). The soil moisture content of the experimental plot at field capacity was measured in the range of 16.18%–18.45%, and the permanent wilting point was in the range of 7.24%–8.56%.

For the ABD trial, the total rainfall during the formative phase of the crop season for the first year (2017) was 86.5 mm and that for the second year was 288.5 mm. Rainfall was unevenly distributed throughout the cropping season in both years, and the crop faced severe moisture stress with a gradual decline in soil moisture content from below 70% to 40% field capacity during the formative growth stages. For the RBD trial, the total rainfall during the formative phase was 101.4 mm in 2019, 170.5 mm in 2020, and 127.3 mm in 2021. Rainfall distribution was uneven in all three years with severe moisture stress and gradual decline in soil moisture content from 60% to 30% field capacity during the crop's formative growth stages. There were intermittent drought cycles with below 50% field capacity during grand growth stages. The control field was given regular irrigation at 100% field capacity until harvest. During the stress period, data on morphological and physiological traits were recorded. At harvest, plant height, tiller numbers, and total biomass per clump were recorded under both control and drought conditions.

Phenotyping during drought stress and field data analysis

During drought, morphological symptoms of leaf drying were recorded through visual scoring. The clones were given a drought score based on the percentage of leaf drying (0%–60%). Physiological data on the leaf relative water content (RWC), canopy temperature, and chlorophyll fluorescence (F_v/F_m) were recorded for the drought-stressed clones and compared with that of control plants. For measuring RWC, approximately 5 cm^2 of fully expanded leaves were cut and placed in pre-weighted tubes and fresh weight (FW) was recorded. The leaves were saturated in distilled water for 5 h at 4°C in darkness and turgid weight (TW) was recorded. Then, the leaves were air-dried in an oven at 65°C for 48 h and dry weight (DW) was recorded. RWC was calculated using the equation of Schonfeld et al. (1988). Canopy temperature was measured using a handheld infrared gun (Model APOGE MI200) between 11:00 a.m. and 1:00 p.m. Chlorophyll fluorescence (F_v/F_m) was measured during clear sunlight days using the chlorophyll fluorometer OS1P (OPTI-SCIENCES, Hudson, USA). The middle portion of the fully expanded second leaf of the main shoot was dark-adapted using the leaf clip and fluorescence was measured. The variable fluorescence (F_v) represents the difference between F_o (the minimum fluorescence) and F_m (the maximum fluorescence). Total chlorophyll content was measured gravimetrically using the method of Arnon (1994). Leaf area (LA) was measured using a non-destructive linear measurement method as mentioned by Montgomery (1911) using the formula LA = LBK (cm^2), where L = maximum length of length, B = maximum breadth, and K = constant (0.75 calculated from the regression analysis). Yield-related traits such as plant height, tiller number, and biomass per clump were recorded at harvest (12 months after planting) for drought-stressed as well as irrigated control plants.

Statistical analysis

The statistical significance of the morphological traits across seasons was evaluated by estimating variance components using a linear mixed model. For the ABD trial, the lme (linear mixed effect model) package in R v4.4.3 was used to calculate the best linear unbiased predictors (BLUPs), and the BLUP values were used for

calculating percent reduction under drought and heritability. For the RBD trial, the restricted maximum likelihood method, implemented in Meta R software, was employed with year as the fixed effect and germplasm and replication as random factors. BLUP-based mean values were subjected to principal component analysis (PCA) using FactoMineR and factoextra packages in R v4.4.3. Broad-sense heritability (%) was estimated in Meta R using the following equation:

$$H^2 = \frac{Vg}{Vg + (Verr/r)}$$

where V_g is the genotypic variance, V_{err} is the error variance, and r is the number of replications.

Stress tolerance and susceptibility index

Stress tolerance index (STI), drought tolerance index (DTI), yield index (YI), and stress susceptibility index (SSI) were computed for each germplasm using the following equations.

$$STI = {}^{(Y_s \times Y_p)}\Big/_{(\bar{Y}_{p^2})} \quad \text{(Fernandez, 1992)}$$

$$DTI = (Y_s \times (Y_s/Y_p))/\bar{Y}_s \quad \text{(Lan, 1998)}$$

$$SSI = (1 - Y_s/Y_p))/(1 - (\bar{Y}_s)/(\bar{Y}_p))) \quad \text{(Fischer and Maurer, 1978)}$$

$$YI = (Y_s)/(\bar{Y}_s) \quad \text{(Guttieri et al., 2001)}$$

where Y_s is the mean value of germplasm under stress conditions and Y_p is the mean value of germplasm under normal conditions.

Selective genotyping by sequencing of the association panel

A selective panel of 96 *Erianthus* germplasm was constructed based on the drought phenotypic data. The panel comprised drought-sensitive and -tolerant clones and clones with intermediate drought reactions. Approximately 1.5 µg of total genomic DNA with an OD_{260}/OD_{280} ratio of 1.8 to 2.0 (assessed using a Qubit 3.0 Fluorometer and NanoDrop 2000 spectrophotometer) from each clone was used for library construction for GBS as described earlier (Gutierrez et al., 2018). Briefly, 0.3–0.6 µg of genomic DNA was digested with *PstI*, ligated with P1 and P2 barcoded adapters followed by PCR amplification. Amplified DNA libraries were quantified using a Qubit® 4.0 Fluorometer, and the insert size was assessed using the Agilent® 2100 Bioanalyzer followed by quantitative real-time polymerase chain reaction (qPCR). Libraries with an appropriate insert size with a concentration of more than 2 nM were pooled and sequenced paired end on an Illumina NovaSeq 6000 platform, with a read length of 144 bp from each end.

Identification of SNPs

Raw sequence reads were filtered using the FastQC method as described earlier and high-quality reads (score >30) were mapped against the *Erianthus rufipilus* (https://sugarcane-genome.cirad.fr/node/1/7) genome (Wang et al., 2023) using bcftools, FreeBayes, and GATK version v4.1 to identify nucleotide variants (SNPs) as described earlier (Shahi et al., 2025). SNPs commonly identified by the three software were filtered at >0.1 minor allele frequency and <50% missing data.

Population structure

To estimate the number of populations, the association panel was assessed using discriminant analysis of principal components (DAPC) with the "adegenet" package of R v4.4.3 (Jombart et al., 2010). Structure analysis was performed with SNPs having high polymorphic information content using the admixture model of STRUCTURE v2.3.4 (Pritchard et al., 2000). The *K* value (i.e., no. of sub-populations) was set at 2–10 with Monte Carlo Markov chain (MCMC) simulation and a burn-in of 20,000 iterations. The corresponding population membership was used as the Q-matrix for marker–trait association analysis. Furthermore, the genetic structure of the lines was determined by cluster analysis based on the shared-allele distance using the neighbor-joining tree algorithm in TASSEL v.5.29 where the branching pattern was assessed with bootstrapping.

Linkage disequilibrium

LD between each pair of markers was estimated as squared allele frequency correlation (r^2) using TASSEL v.5.29 (Bradbury et al., 2007). LD decay distance was estimated by plotting the r^2 values between marker pairs against physical distance. The intersection points of the LOESS curve and the r^2 threshold in the scatter plot was identified as the average LD decay across the genome.

Genome-wide association analysis

A genome-wide association study (GWAS) was conducted using the best linear unbiased estimates (BLUEs) of the phenotypic data collected across 3 years (2021–2023) of a panel comprising 96 *Erianthus* clones and 1,044 high-quality SNPs to identify QTNs associated with morpho-physiological traits under drought stress. GWA analysis was conducted in the R package GAPIT v.3 (Wang and Zhang, 2021) using the multi-locus models such as fixed and random model circulating probability unification (FarmCPU), Bayesian-information and linkage-disequilibrium iteratively nested keyway (BLINK), and multiple loci mixed model (MLMM) with default parameters. False discovery rate (FDR) < 5% (Benjamini and Hochberg, 1995) was used as the threshold to determine significant QTNs. Furthermore, six other multi-locus GWAS models, such as

mrMLM, FASTmrMLM, FASTmrEMMA, pLARmEB, ISIS EM-BLASSO, and pkWmEB, were implemented in the R package mrMLM v5.0 (https://cran.r-project.org/web/packages/mrMLM/index.html) with default parameters. The QTNs with LOD score ≥3.00 were declared significant. A QTN identified in at least two models was designated as a consistent QTN, and a QTN observed in three or more models with >10% phenotypic variation explained (PVE or R^2) at a p-value < 0.00001 was considered as a major QTN.

Identification of candidate genes

Genes within 200-kb flanking regions of the significant QTNs were retrieved from the *E. rufipilus* reference genome (https://sugarcane-genome.cirad.fr/node/1/7). Downstream gene ontology (GO) enrichment analysis was performed using two gene annotation databases DAVID (https://davidbioinformatics.nih.gov) and agriGO V2.0 (http://systemsbiology.cau.edu.cn/agriGOv2/).

Results

Comparative drought performance of *Erianthus* clones

During moisture stress at below 60% field capacity in the ABD trial, leaf drying followed by leaf rolling were the first symptoms observed in a few *Erianthus* clones. At <50% field capacity, the clones were rated from tolerant to susceptible based on the clearly visible leaf drying symptom on the scale of 0%–60% (Figures 1a, b). At soil moisture below 40% field capacity, 18 clones were scored as highly tolerant (no leaf drying) or tolerant (10% leaf drying), 20% for moderately tolerant clones, 37 clones with 20% as moderately tolerant, 32 with 30%–40% as moderately susceptible, 39 clones with more than 50% leaf drying as susceptible, and 34 clones were highly susceptible with 60% leaf drying (Figures 1c, d). The mean percentage increase in canopy temperature and the percentage reduction in chlorophyll fluorescence (F_v/F_m) and total chlorophyll content under drought stress compared to control conditions corresponded to the leaf drying scores (Figures 2a–c; Supplementary Figure 1).

The increase in canopy temperature was 2%–15% lower among the highly tolerant and tolerant clones, while the clones in the intermediate and susceptible category showed a 20%–25% and 28%–47% increase, respectively. Canopy temperature was in the range of 23–24°C in control as well as highly tolerant clones under drought conditions while the susceptible clones recorded higher (29–35°C) canopy temperature. The mean reduction in chlorophyll fluorescence among the tolerant clones ranged from 8% to 25%, while the susceptible clones showed a 30%–60% reduction. A similar trend was observed for total chlorophyll content where the tolerant clones maintained a high chlorophyll content with minimum reduction at 6.25%–12%. The physiological performance of the tolerant clones was reflected in the final biomass data, with minimum reduction in plant height and biomass (Figures 2d, e). Highly tolerant and tolerant clones did not show any reduction in plant height, while the moderately tolerant and susceptible clones showed 15%–22% reduction compared to that of control. Mean biomass per clump ranged from 83.25 to 9.55 kg with a percentage reduction ranging from 5% to 16% among the tolerant clones and 30% to 52% among the susceptible clones.

PCA was performed based on mean data obtained from BLUP across 2 years' data on canopy temperature (CTC), total chlorophyll content (TCC), chlorophyll fluorescence (CFC), plant height (PHC), and biomass (BMC) in 223 *Erianthus* accessions. The distribution and associations between various traits are depicted in a biplot against the first two principal components PC1 and PC2 (Figures 3a, b). The first two principal components collectively explained 52.9% of the overall variability under control conditions compared to 70.18% under drought conditions, highlighting a more concentrated distribution of variation under drought conditions. The distribution of accessions was relatively well-spread under control conditions (Figure 3a), while under drought conditions, it was more compressed (Figure 3b), and trait expression was influenced mostly by the first dimension. The traits were grouped into three major vectors under control conditions and into two groups under drought conditions. Under both conditions, chlorophyll fluorescence, biomass, and plant height showed positive association and more contribution to the first principal component. Under control conditions, canopy temperature and chlorophyll fluorescence primarily contributed more to the second principal component, whereas under drought conditions, only canopy temperature did so. The PCA biplot of 223 *Erianthus* germplasms clearly clustered tolerant and susceptible genotypes. The broad-sense heritability (H^2) estimates were generally higher and more for all the traits except chlorophyll fluorescence under drought conditions (0.60–0.99) compared to the control (0.39–0.96) (Figure 3c).

Phenotypic performance of the drought association panel

The association panel consisting of 91 accessions with varying drought responses was drought-stressed during the formative phase where the respective controls were irrigated regularly. Approximately 23 clones were highly tolerant with 0%–10% leaf drying, 27 clones were tolerant with 20%–30%, 23 clones with 40%–50%, and 18 clones were highly susceptible, showing more than 60% leaf drying. The clones showed significant differences for canopy temperature, chlorophyll fluorescence (F_v/F_m), RWC, and total chlorophyll content under drought stress where the mean difference showed reduction compared to control conditions (Figure 4; Supplementary Figure 2). However, the reduction was significantly lower among the tolerant clones and higher in the susceptible clones. Highly tolerant clones maintained low canopy leaf temperature in the range of 23–26°C with a mean increase from 6% to 12%, while most susceptible clones recorded a leaf temperature of 30–32°C with a mean increase of 24%–28% under drought stress (Figure 4a; Supplementary Figure 2). The tolerant clones maintained high chlorophyll fluorescence (F_v/F_m) with 9%–

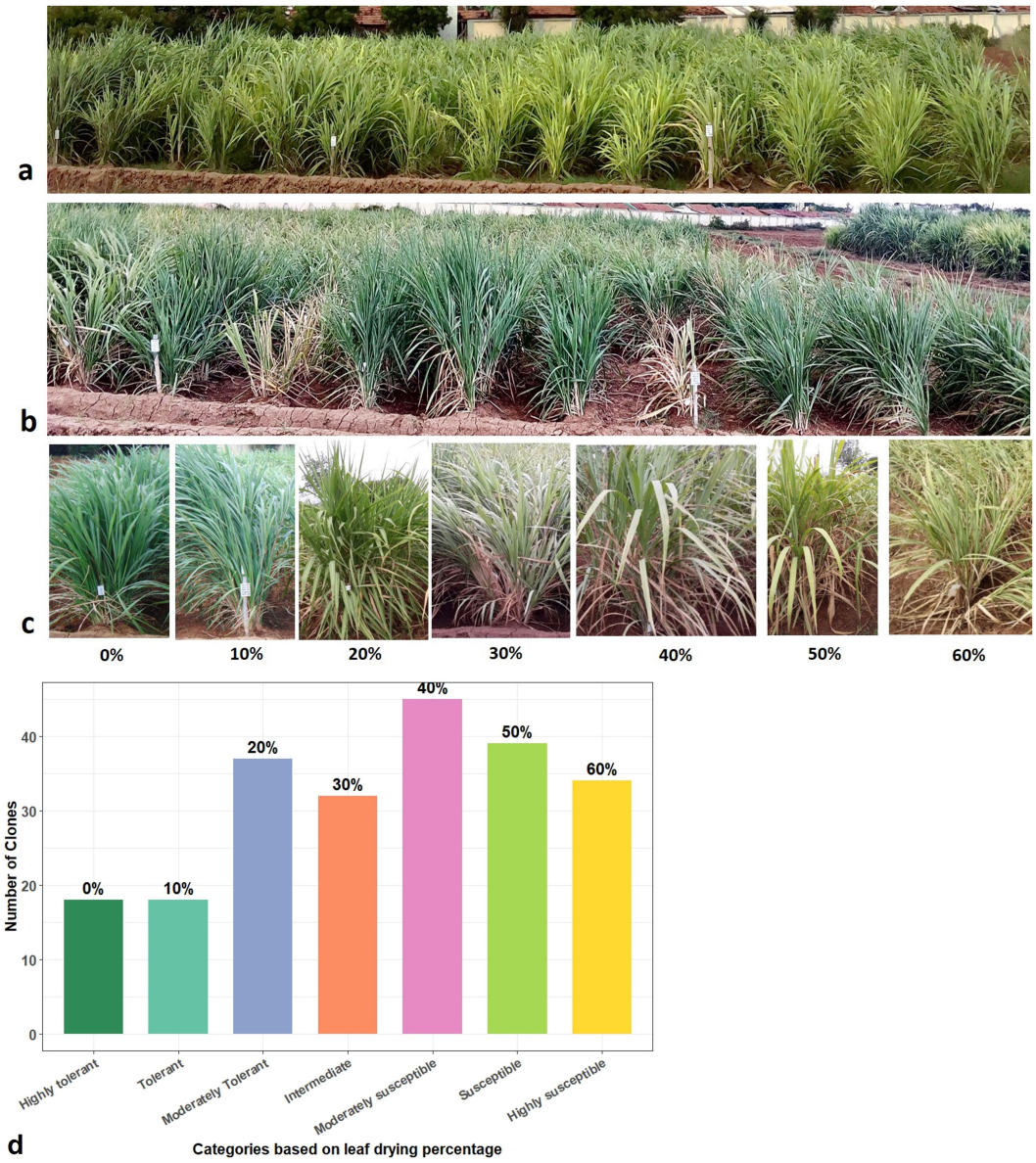


FIGURE 1
Field evaluation of *Erianthus* clones for drought tolerance response. **(a)** *Erianthus* clones under irrigated control conditions. **(b)** Erianthus clones showing leaf drying symptom under drought stress conditions. **(c)** Leaf drying percentage and morphology. **(d)** Distribution of germplasm accessions grouped based on leaf drying.

18% mean reduction, while the susceptible clones showed more than 20% reduction (Figure 4b). Similarly, highly tolerant and tolerant clones showed lower mean reduction in leaf RWC (6%–18%), tiller number (7%–18%), and LA (5%–19%) under drought conditions (Figures 4c–e). The physiological performance of clones corresponded with the mean reduction in biomass yield at harvest. The tolerant clones showed 18%–22% less biomass whereas the susceptible clones had more than 42% less final biomass under drought conditions (Figure 4). Among the highly tolerant clones, five clones (SES 133, SES 149, SES 288, SES 293 and SES 347) consistently exhibited drought tolerance by maintaining higher chlorophyll fluorescence, RWC, leaf area, lower canopy temperature and increased biomass accumulation combining both the trials for five-years.

The PCA revealed a distinct pattern of variation in the association panel under control and drought conditions (Figures 5a, b). The first two principal components explained 57.2% of the total variation under control conditions with an eigen value of more than one. The accessions were widely spread, indicating wide variation in trait expression. LA, biomass, and tiller number were positively associated traits with a significant contribution to total variation. Under drought conditions, the first principal component alone explained 67.4% out of the 80.3% variation captured by the first two principal components with an

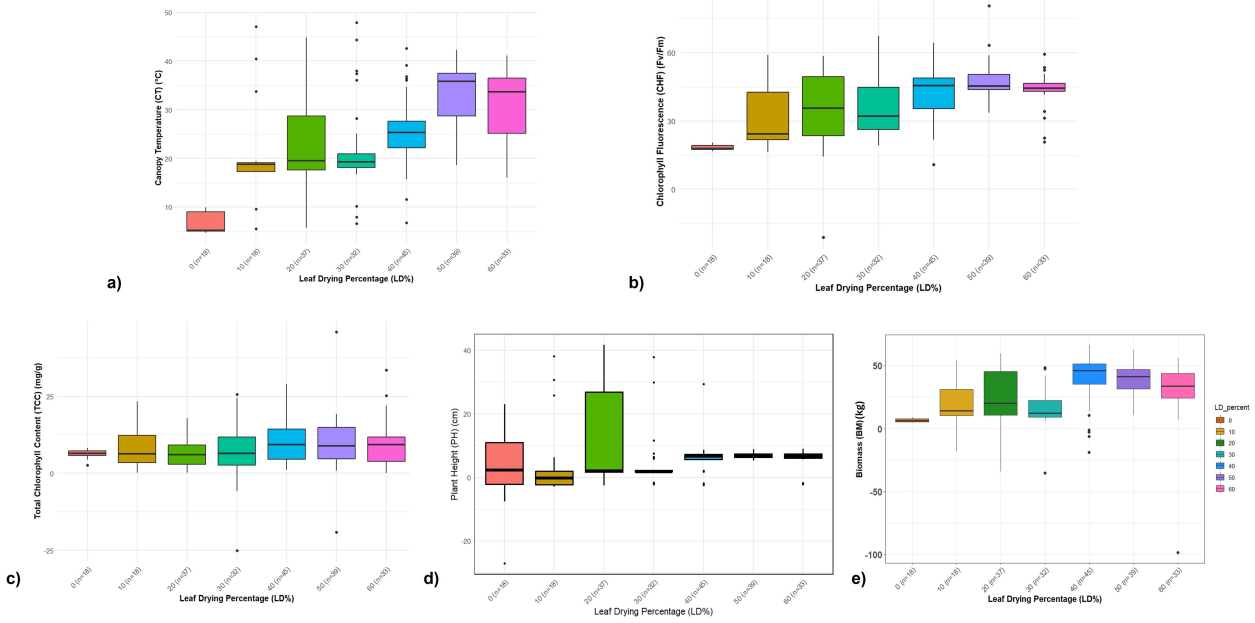


FIGURE 2
Categorization of germplasm accession based on drought tolerance and trait analysis of germplasm based on leaf drying percentage for the 2-years ABD trial. **(a–e)** Boxplot showing the relationship between different physiological traits among *Erianthus* clones categorized based on leaf drying percentage under drought stress. **(a)** Percent increase in canopy temperature. **(b)** Percent reduction in chlorophyll fluorescence (F_v/F_m). **(c)** Percent reduction in total chlorophyll content. **(d)** Percent reduction in plant height. **(e)** Percent reduction in biomass among *Erianthus* clones under drought stress relative to irrigated control conditions.

eigen value of more than one. Majority of the traits including canopy temperature, chlorophyll fluorescence, LA, RWC, and biomass showed positive correlation with a greater contribution to the first principal component. Canopy temperature was the sole significant contributor to the second principal component under drought conditions. PCA clearly clustered the tolerant and susceptible clones under both control and drought stress conditions. The heritability percent estimated for various physiological traits were in the range of 74%–96% under control conditions and in the range of 86%–99% under drought conditions; tiller number showed the maximum heritability percentage under both control and drought conditions (39-96) (Figure 5c).

Stress tolerance and susceptibility indices of clones in the association panel

The performance of *Erianthus* clones in the association panel with respect to final biomass under drought stress showed that the STI ranged from 0.07 to 2.50 with a grand mean of 0.79 (Figure 6a) with 24 (highly) tolerant clones having higher STI (>1.01) for biomass compared to the susceptible clones. The DTI was in the range of 0.08 and 1.76 (Figure 6b) where the tolerant clones showed higher DTI (0.82 to 1.21). A total of 23 clones had a higher DTI (>1.02) while 3 clones (IND 99-892, IND 10-1591, and IND-03 1260) recorded a higher DTI than the tolerant check. The average YI calculated based on the biomass of the clones was 1.0 with a range from 0.22 to 1.98 (Figure 6c). Ten germplasms displayed higher YI at values more than 1.6. SSI ranged from 0.17 to 2.48 with an average of 1.02. SSI for all highly tolerant clones was less than 1.0 (Figure 6d). Two *Erianthus* clones (SES 288 and SES 293) with superior physiological performance recorded higher STI (1.92 and 1.87), DTI (1.39 and 1.36), YI (1.69 and 1.66) and lower SSI (0.69 and 0.70).

Genotyping by sequencing and variant calling

GBS of 91 *Erianthus* germplasm clones generated 5.143 Gb raw data and 5.142 Gb clean data after filtering out low-quality data. The average raw data generated for the clones were 1.83 million reads, which, after QC, resulted in 1.76 million reads with a Phred score of more than 20. Mapping of cleaned short reads against the *E. rufipilus* (https://sugarcane-genome.cirad.fr/node/1/7) reference genome identified 3,696 bi-allelic SNPs common between three software. Further filtering with a minor allele frequency set at 5% and 20% missing genotype resulted in 1,044 SNPs that were used for downstream marker–trait association.

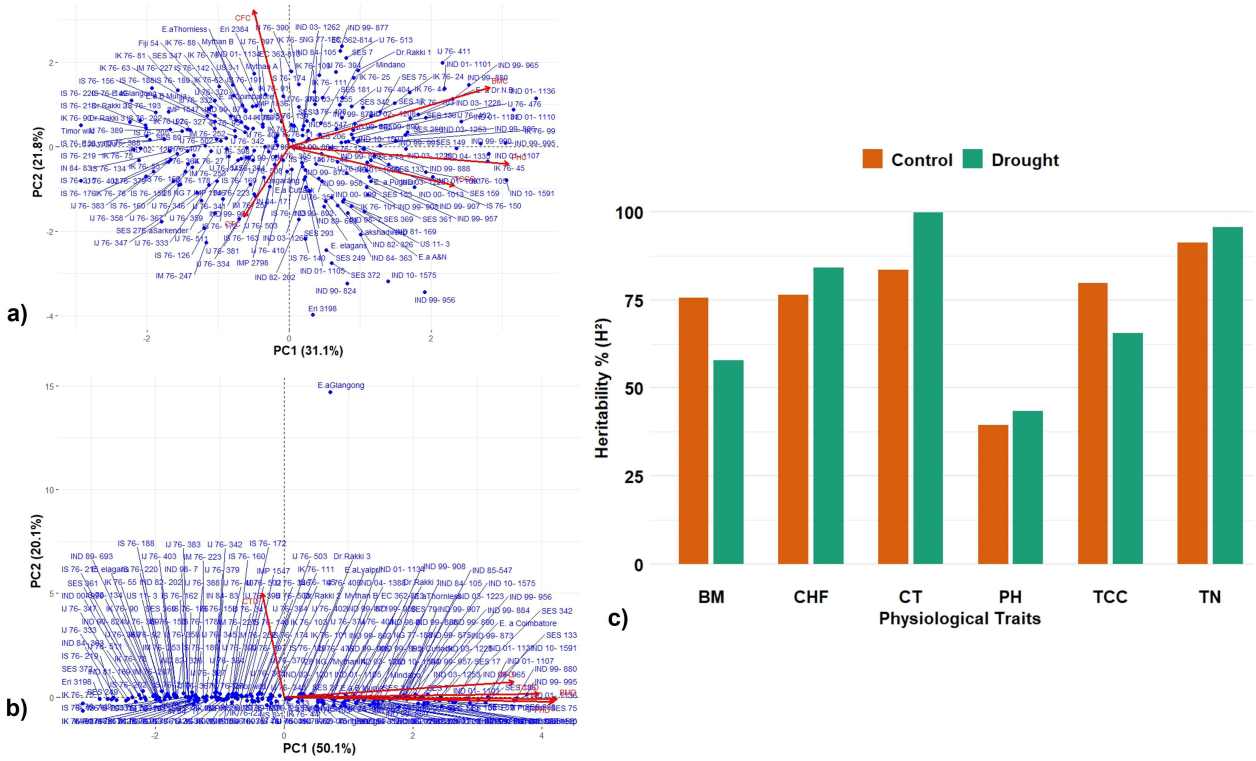


FIGURE 3
Principal component analysis (PCA) of germplasm accessions and trait relationship and heritability index of physiological traits in *Erianthus* clones under control and drought conditions. Biplot of the first two principal components of 223 germplasm accessions among the studied traits under control **(a)** and drought **(b)**. PC1 (Dimension 1), PC2 (Dimension 2). Broad sense heritability % (H^2) of different physiological traits under control and drought conditions **(c)**.

Population structure and linkage disequilibrium

Population structure analysis, based on the Δ*K* method, categorized 91 genotypes into two sub-populations, SP1 and SP2, consisting of 43 and 48 genotypes, respectively. Individuals of each population were further categorized into two groups, i.e., pure and admixture types, where populations comprising ≥0.8 of the member proportions were considered pure and others as admixtures. Based on this criterion, SP1 was 75% pure and 25% admixtures and SP2 was 63% pure and 37% admixtures (Figures 7a, b). Similar results were found by Tsuruta et al. (2022) where the high values of Δ*K* were obtained at $K = 2$ and $K = 3$ with 121 *Erianthus* accessions from Thailand based on the marker profiles with 28 SSR primers. The tree based on the neighbor-joining method also identified two clusters (Figure 7c), indicating consistency in the grouping of genotypes. From the PCA, PC1 explained 32.8% of the genetic variance while PC2 explained 13.0% (Figure 7d), which distinguished the two sub-populations based on their geographic regions. The background LD in the analyzed association mapping panel (AM) was equal to 0.2, which was taken as the threshold cutoff for estimating LD decay. The whole-genome LD in the AM panel decayed at 3 Mb (Figure 7e).

Genome-wide association analysis

Genome-wide association mapping using six MrMLM models identified 43 QTNs distributed over 10 chromosomes associated with 12 traits under control and drought conditions. Among the six models, ISIS EM-BLASSO and pLARmEB revealed the maximum number of QTNs, i.e., 30 and 22, respectively, whereas FASTmrEMMA detected the lowest number (10) of QTNs. Similarly, 46 QTNs in 10 chromosomal regions were detected using four models in GAPIT. Of these QTNs, 23 were common in at least two models (Table 1; Figure 8). The phenotypic variation explained (PVE or R^2) by the QTNs ranged from 0.26% to 59.35%, indicating that traits under both control and drought conditions were controlled by multiple loci with small to moderate effects. The most significant QTN was recorded for biomass under drought conditions (BM_DT) linked to marker *S02_5349882* at $-\text{Log}_{10}$ *p*-value = 6.2 and R^2 = 54.1% followed by tiller number under drought (TN_DT) linked to *S04_9908350* at –Log10 *p*-value = 6.9 and R^2 = 20.1%. The highest number of QTNs was identified on Chr2 (22) followed by Chr6 (17) (Table 1; Figure 8). Three major QTNs (*S02_5349882* identified by five models, *S03_16447857* by four models, and *S04_9908350* by five models) were identified in Chr2, Chr3, and Chr4, respectively. Trait-wise, QTNs ranged from

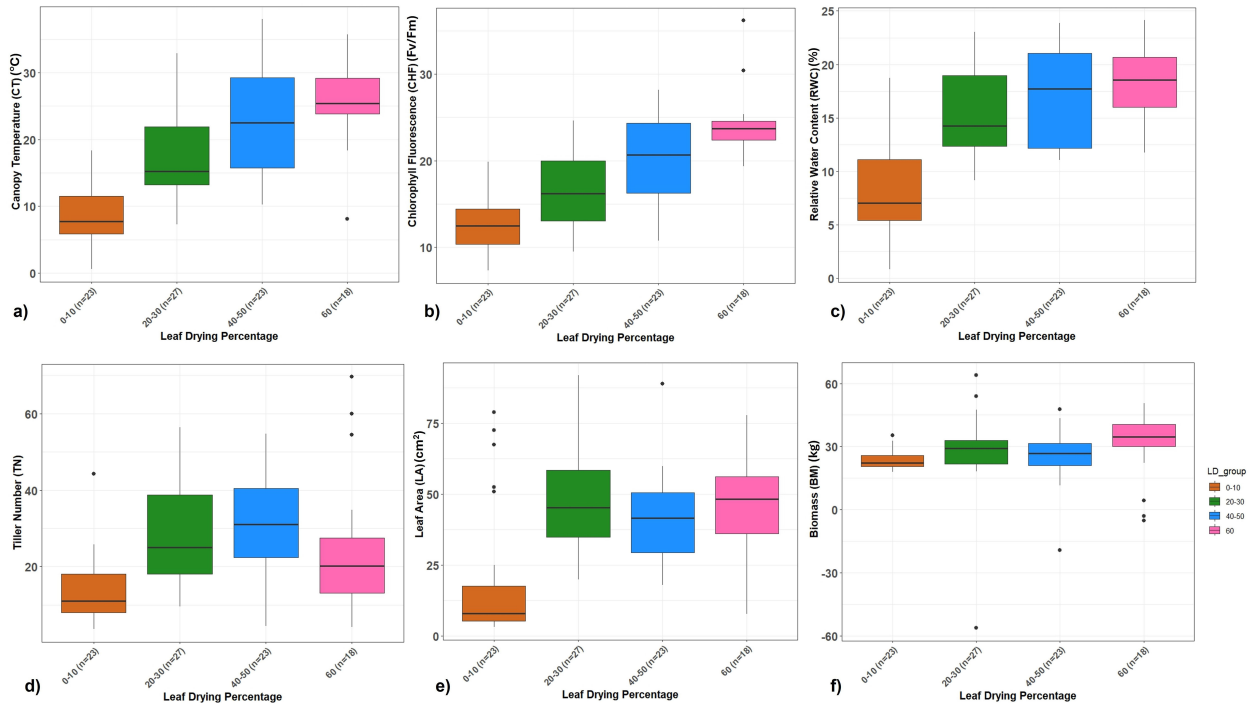


FIGURE 4
(a–f) Boxplot showing the relationship between different traits vs. leaf drying percentage for the 3-year RBD trial. **(a)** Percent increase in canopy temperature. **(b)** Percent reduction in chlorophyll fluorescence (F_v/F_m). **(c)** Percent reduction in relative water content. **(d)** Percent reduction in tiller number. **(e)** Percent reduction in leaf area. **(f)** Percent reduction in biomass of *Erianthus* clones under drought stress relative to irrigated control conditions.

one each for RWC under control (RWC_C) and LA under drought (LA_DT) conditions to seven QTNs associated with tiller number (TN_DT) as well as canopy temperature (CT_DT) under drought conditions (Table 1). Altogether, 26 QTNs were identified to be associated with different traits under drought conditions whereas 17 QTNs were identified for traits under control conditions. Seven QTNs were common between control and drought conditions, while 9 and 17 QTNs expressed specifically under control and drought conditions, respectively. One QTN (*S03_16447857*) was associated with both LA (LA_C) and tiller number (TN_C) under control conditions. Moreover, two QTNs (*S05_47721536* and *S08_2099940*, both with FDR *p*-value = 0.013) identified by MrMLM as well as the GAPIT model were associated with TN_DT. On the other hand, one QTN (*S08_2099940*) showed association with more than one trait (CT_DT, RWC_DT, and TN_DT) under drought stress (Table 1).

Candidate genes identified in the QTNs

A total of 235 genes were identified linked to 34 QTNs associated with different traits under drought stress, of which 34 genes were unique and 27 have previously known functional annotations (Supplementary Tables 1, 2). Maximum candidate genes were identified from significant QTNs associated with different traits under drought stress. QTN (*S08_2099940*) associated with three physiological traits under drought conditions (CT_DT, TN_DT, and RWC_DT) was annotated as a regulatory-associated protein of TOR 2 (Target of Rapamycin 2). Several candidate genes such as potassium transporter (QTN-*S06_37343127*), auxin response factor (QTN-*S05_55750756*), FAD-binding PCMH-type domain-containing protein (QTN-*S03_493324*), protein kinase domain-containing protein (QTN-*S05_74779768*), and EG45-like domain-containing protein (QTN-*S10_29330993*) were identified from QTNs associated with chlorophyll fluorescence under drought conditions (CF_DT). Two candidate genes encoding receptor protein kinase TMK1 (Trans-Membrane Kinase 1) (QTN-*S01_17733326*) and putative NRT1/PTR (Nitrate Transporter 1/Peptide Transporter Family) family protein (QTN-*S10_47224375*) were identified to be associated with TN_DT. Two candidate genes such as putative pentatricopeptide repeat-containing protein and putative polyubiquitin were identified from QTNs linked to markers *S02_19373704* and *S10_62906516*, respectively, for RWC_DT.

Discussion

Sugarcane achieves 70%–80% of yield during its formative phase while the remaining yield is attained during the grand growth phase. Drought stress during these critical growth stages with high water requirement (Hemaprabha et al., 2013; Reyes et al., 2021) severely affects cane growth, yield, and sucrose content (Silva

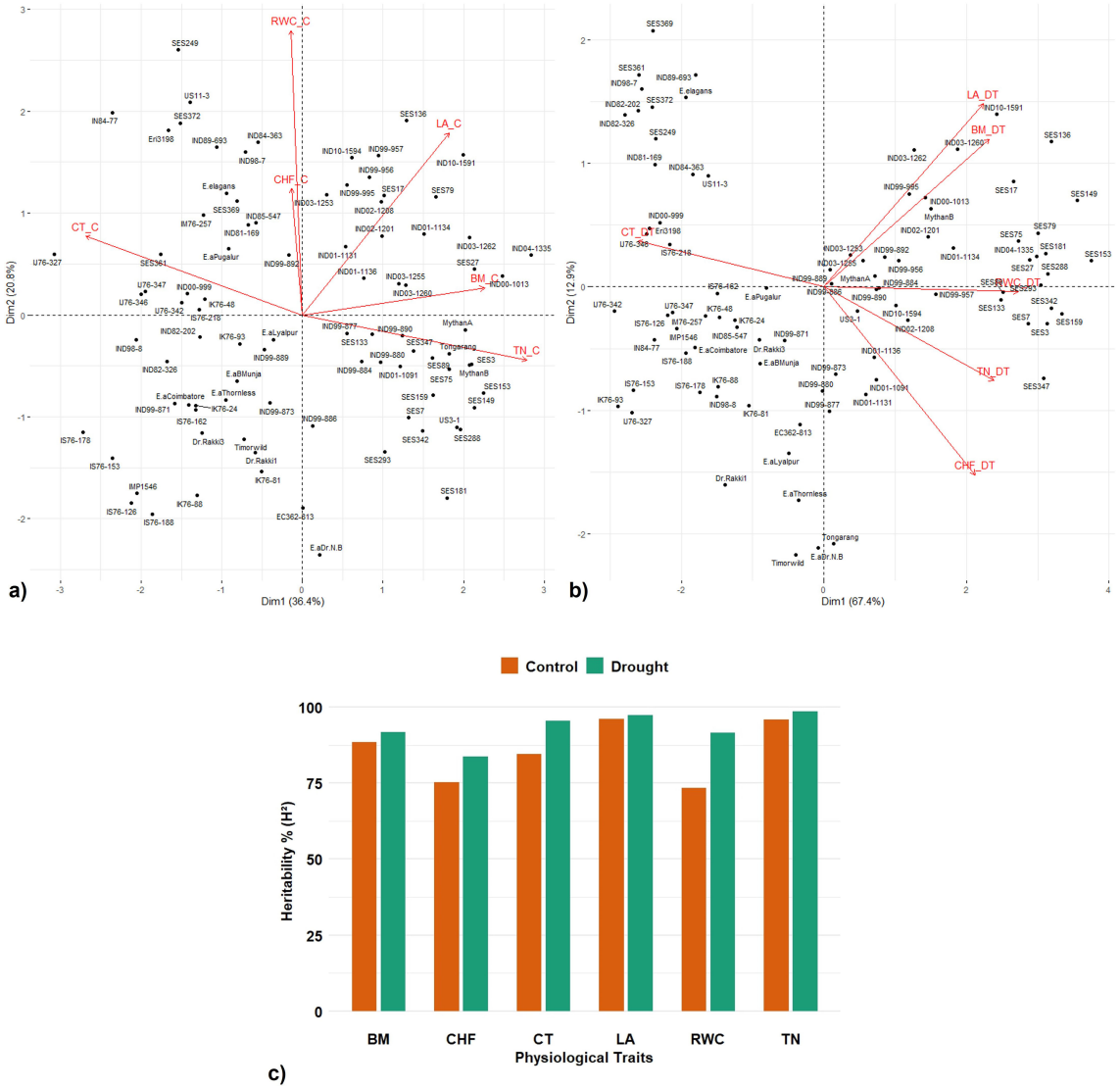


FIGURE 5
Principal component analysis (PCA) of germplasm accessions and trait relationship and heritability index under control and drought conditions. Biplot of the first two principal components of 91 accessions among the traits under control **(a)** and drought **(b)**. PC1 (Dimension 1), PC2 (Dimension 2). **(c)** Comparative broad sense heritability estimates of various traits under control and drought conditions. TN—tiller number, CF—chlorophyll fluorescence, CT—canopy temperature, LA—leaf area, RWC—relative water content, and BM—Biomass.

et al., 2007; Ferreira et al., 2017) with up to 50%–60% yield reduction depending on the severity and number of drought spells (Vinoth et al., 2022). Most of the commercial sugarcane genotypes with a narrow genetic base show poor tolerance to drought (Hemaprabha et al., 2013; Mohanraj et al., 2021). Therefore, there is a need for alternative sources of resistance to facilitate the development of sugarcane genotypes that can not only withstand the extreme environmental stress but also sustain yield. *Erianthus* has been identified as a sugarcane-related genus with many significant drought-adaptive morphological and physiological traits (Valarmathi et al., 2023). Extensive phenotyping and identification of the sequence variations in *Erianthus* spp. controlling critical agronomic traits under drought conditions could help identify resources to improve sugarcane productivity in drought-prone areas.

Higher photosynthetic efficiency among tolerant *Erianthus* clones

Erianthus clones employ several important physiological and biochemical traits to adapt and mitigate the effect of water stress (Valarmathi et al., 2023). Maintaining higher photosynthetic efficiency, cooler canopy, and higher RWC; accumulating osmotic solutes; and enhancing water uptake through its deep root system, *Erianthus* clones thrive under water stress with high water use

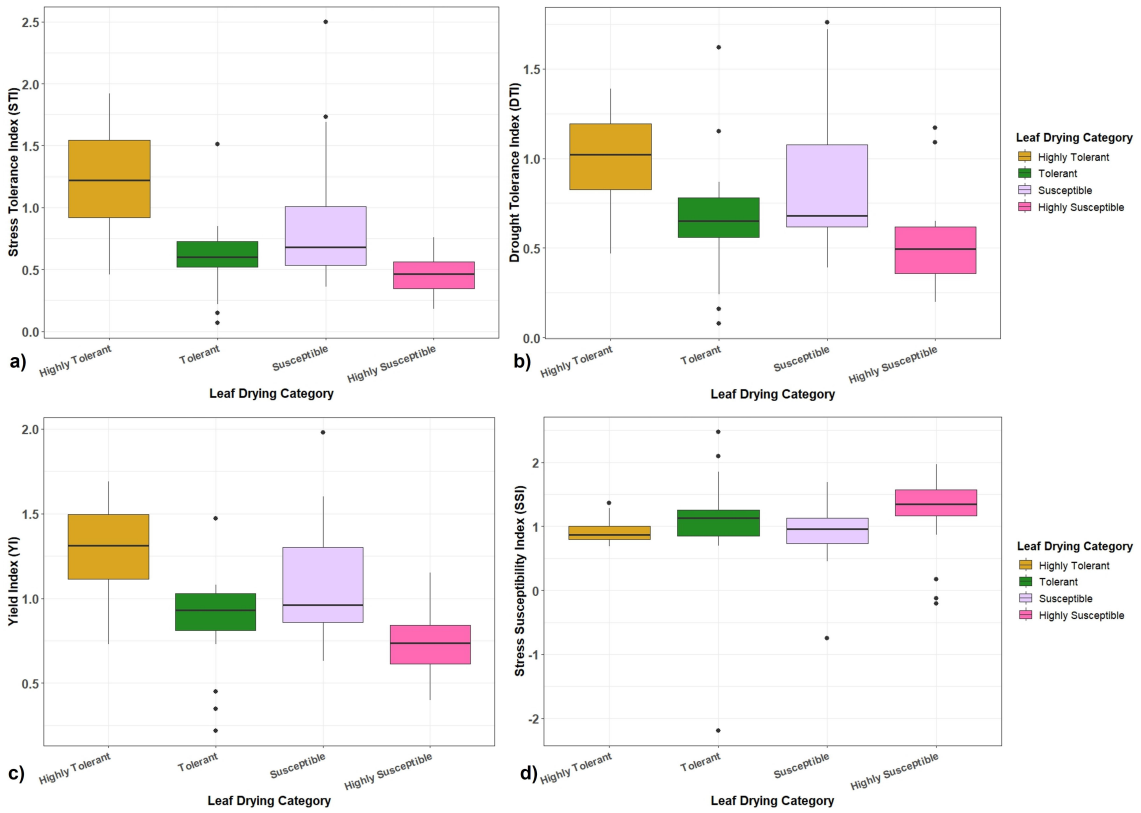


FIGURE 6
Boxplots showing the relationship between different drought tolerance indices with highly tolerant, tolerant, susceptible, and highly susceptible categories of *Erianthus* clones. **(a)** Stress tolerance index (STI). **(b)** Drought tolerance index (DTI). **(c)** Yield index (YI). **(d)** Stress susceptible index (SSI).

efficiency to sustain plant growth and biomass production (Baby et al., 2005). Superior physiological and growth responses of clones at the formative stage of drought stress can be early indicators to select drought-tolerant genotypes with a better yield.

Drought stress impedes plant growth primarily through the reduction in photosynthetic efficiency, which serves as a reliable prime physiological biomarker to assess drought stress response (Wang et al., 2025). Intact green canopy maintained in the highly tolerant and tolerant *Erianthus* clones shows less photosynthetic organ damage, leading to the maintenance of high photosynthetic efficiency during drought stress. Canopy leaf temperature and chlorophyll fluorescence (F_v/F_m) are other important physiological traits to sustain photosynthetic activity under drought stress. Normal transpiration maintains the canopy temperature at a metabolically functional range, while closure of stomata during high evaporative demands under drought stress leads to a high canopy temperature (Siddique et al., 2000). The lower canopy temperature recorded among tolerant genotypes could be due to the efficient extraction of available soil moisture by their deep root system to maintain normal transpiration and transpirational cooling (Reynolds et al., 2010). Chlorophyll fluorescence (F_v/F_m), an important photochemical quenching parameter that indicates the quantum efficiency of PSII and reduction in F_v/F_m, suggests downregulation of photosynthesis due to the inactivation of PSII activity (Yao et al., 2018). Along with chlorophyll fluorescence, reduction in RWC and chlorophyll content is also associated with decreased photosynthetic efficiency (Kumar et al., 2024). The tolerant *Erianthus* genotypes with a higher chlorophyll fluorescence, RWC, and chlorophyll content were able to regulate photosynthetic rates as well as transpirational cooling under water stress conditions, which supported growth and final biomass accumulation.

Higher drought tolerance indices recorded for tolerant *Erianthus* clones

Stress tolerance indices provide a measure of drought effects to identify drought-tolerant genotypes where genotypes with higher drought indices are more likely to be tolerant and stable under stress conditions (Wirojsirasak et al., 2023). Different drought tolerance indices across *Erianthus* germplasms revealed significant differences in their adaptability under drought stress to sustain biomass production. Highly tolerant and tolerant clones showing consistently higher tolerance indices (STI and DI) and lower SSI indicate the ability of these clones to sustain growth and biomass

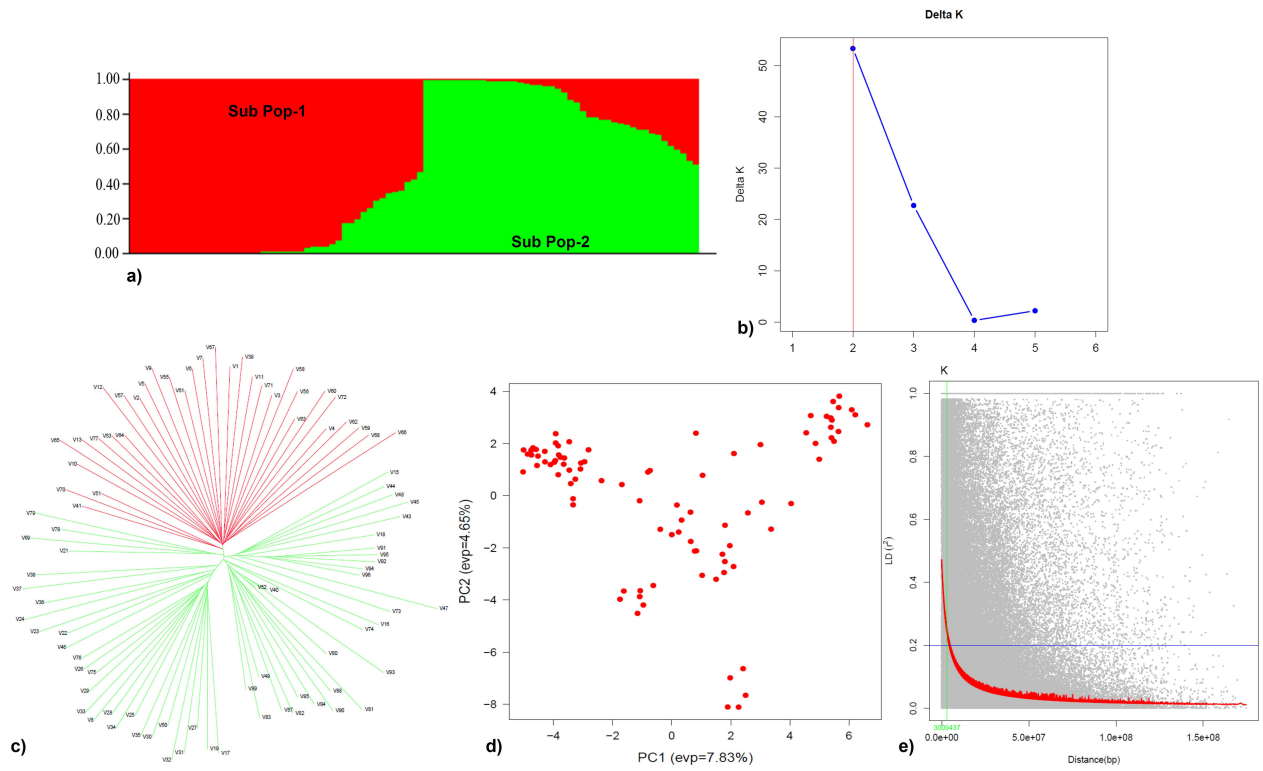


FIGURE 7
Population structure, principal component, and linkage disequilibrium analyses of the *Erianthus* drought panel. **(a)** Bar plot representing two subpopulations (Sub_pop 1 and Sub_pop 2). **(b)** Delta K plot—Estimation of optimum number for subpopulation (highest peak at *K* = 2). **(c)** Neighbor-joining tree of 91 germplasm accessions illustrating the separation of studied germplasm accessions into two main clusters corresponding to structure analysis. **(d)** Scatter plot of the first two principal components (PC1 and PC2) showing variation present in the accessions. **(e)** Scatter plot showing the LD decay with physical distance (bp); red color indicates the fitted LD decay.

TABLE 1 Quantitative trait nucleotides (QTNs) associated with various traits under control and drought conditions in the *Erianthus* drought panel.

Traits	QTNs	Chr	Position (bp)	Model	P value	PVE/R^2 (%)	LOD score	Effect	MAF
BM_C	S02_5349882	2	5349882	1,2,4,6, 9	3.9-4	19.9-45.7	3.27-3.29	3.2	0.13
BM_C	S03_9598166	3	9598166	5	4.3	19.5	3.58	-8.7	0.13
BM_C	S06_13102979	6	13102979	5	5.4	31.8	4.66	10.1	0.31
BM_DT	S02_5349882	2	5349882	1,2,3,4,5	5.3-6.2	35.3-54.13	4.6-5.4	14.1	0.13
BM_DT	S03_9598166	3	9598166	4	5.01	23.12	4.25	-9.2	0.13
BM_DT	S06_13102979	6	13102979	4	5.65	29.21	4.86	9.4	0.31
CT_C	S02_5349863	2	5349863	1,2,3,4,5,6	3.9-6.7	21.4-59.3	3.2-5.9	2.8	0.23
CT_C	S01_9807671	1	9807671	3,5	3.7-4.5	17.7	3-3.7	0.9	0.12
CT_DT	S02_5349863	2	5349863	1,2	3.8-4.3	21.1-24.6	3.1-3.5	2.4	0.23
CT_DT	S02_19373707	2	19373707	1,2	3.8-4.9	16.4-18.5	3.1-4.2	2.1	0.25
CT_DT	S08_2099940	8	2099940	1,2,3,4,5,7,8,9,10	4.1-6.7	12.8-45.5	3.4-5.9	-1.8	0.14
CT_DT	S05_55750974	5	55750974	4,5	6.7-7.5	14.7-21.9	5.9-6.6	-1.7	0.39
CT_DT	S01_9807671	1	9807671	5	5.6	11.1	4.8	1.7	0.12
CT_DT	S09_15889790	9	15889790	5	6.482	20.8	5.7	-2.1	0.20
CT_DT	S09_27376448	9	27376448	5	4.3919	14.8	3.7	1.8	0.27

(Continued)

TABLE 1 Continued

Traits	QTNs	Chr	Position (bp)	Model	P value	PVE/R^2 (%)	LOD score	Effect	MAF
CHF_C	S01_102019624	1	102019624	1,3	3.8-4.5	35.9	3.1-3.7	0.02	0.24
CHF_C	S05_79251361	5	79251361	5	3.8-4.5	17.8-25	3.1-3.7	1.00E-02	0.18
CHF_C	S02_50417983	2	50417983	1,2	4.8409	10.02	4.0852	0.008	0.24
CHF_DT	S06_37343127	6	37343127	6	5.0071	49.46	4.2438	-0.042	0.31
CHF_DT	S03_493324	3	493324	5	6.3419	7.73	5.5256	0.018	0.25
CHF_DT	S05_55750756	5	55750756	5	6.314	28.01	5.4987	0.0351	0.31
CHF_DT	S05_74779768	5	74779768	5	6.3216	1.22	5.506	0.0084	0.12
CHF_DT	S10_29330993	10	29330993	5	6.0604	0.26	5.2543	0.0035	0.21
LA_C	S03_37378020	3	37378020	1,2,4,5,6	4.5-5	23.8-43.4	3.7-4.2	-60.1582	0.13
LA_C	S05_98492320	5	98492320	2,4,9	4.04	0.5-0.7	3.32	39.5	0.23
LA_DT	S09_66579665	9	66579665	1,2,4,5,6	4.2	20.1-46.6	3.5	70.34	0.34
RWC_C	S06_45836308	6	45836308	1,2,3,4,5,7,8,9,10	3.7-5.1	7-27.4	3.03-4.4	-1.0253	0.46
RWC_DT	S08_2099940	8	2099940	1,3,4,6	3.9-5	16.9-42.2	3.2-4.2	4.2	0.14
RWC_DT	S02_19373704	2	19373704	4	3.8	10.4	3.1	2.4	0.19
RWC_DT	S10_62906516	10	62906516	5	4.3	9.6	3.6	1.7	0.48
TN_C	S03_16447857	3	16447857	1,2,4,5	5.7-5.9	9.4-16.5	4.9-5.1	5.1	0.42
TN_C	S04_16818232	4	16818232	1,2,4,5,6	4.2-5.8	15.4-25.4	3.1-5	6.3	0.25
TN_C	S08_2099940	8	2099940	1,2,3,4,5,6	3.7-4.9	13.5-36	3-4.2	7.1	0.14
TN_C	S04_37730964	4	37730964	4,5	3.8-4.7	5.8-10.1	3.1-4	4.7	0.42
TN_C	S06_31437384	6	31437384	4,5, 9,10	4.1-4.8	9.8-9.9	3.4-4.1	-3.9	0.20
TN_C	S02_89391771	2	89391771	6	6.6429	20.54	5.8162	-6.7	0.36
TN_DT	S04_9908350	4	9908350	1,2,4,5,6	5.1-6.9	16.4-20.1	4.3-6.1	6.3	0.34
TN_DT	S05_32454167	5	32454167	1,2,3,4,5,6	3.8-4.8	18.3-42.9	3.1-4.1	-5.6	0.18
TN_DT	S03_16447857	3	16447857	3	4.4	17.9	3.6	10.4	0.41
TN_DT	S01_17733326	1	17733326	4,5	3.8	6.5	3.1		0.26
TN_DT	S05_47721536	5	47721536	4,5,7,8,9,10	4.3	10.5	3.6	-4.5	0.15
TN_DT	S08_2099940	8	2099940	4,5	3.8	11.7-11.9	3.1	5.0	0.14
TN_DT	S10_47224375	10	47224375	4,5	3.9	4.9	3.2	-4.8	0.45

1-mrMLM, 2- FASTmrMLM, 3-FASTmrEMMA, 4-pLARmEB, 5-ISIS EM-BLASSO,6-pKWmEB,7-MLMM,8-FarmCPU,9-BLINK, 10-CMLM; bp, base pair; BM, biomass; Chr, chromosome; CT, canopy temperature; CHF, chlorophyll fluorescence; LA, leaf area; MAF, minor allele frequency; PVE/R^2, phenotypic variance explained; RWC, relative water content; TN, tiller number.

production with minimal reduction under drought stress (Augustine et al., 2015; Sanghera and Kashyap, 2022). The susceptible genotypes with a high SSI and a poor STI score supported their significant stress-induced biomass reduction.

Marker–trait association and linked genes

Genome-wide association analysis detected a total of 43 QTNs distributed on 10 different chromosomal regions, which explained between 0.26% and 59.35% of the total phenotypic variation, indicating that the traits were controlled by multiple small–moderate effect loci. The 17 QTNs that were uniquely identified under drought stress are important for consideration of further validation using the genotypes that were not used for this GWAS. Of special importance are the two significant QTNs (S05_47721536 and S08_2099940) for the tiller number under drought conditions, specifically the pleiotropic QTN (S08_2099940) that was significantly associated with more multiple traits such as canopy temperature, RWC, and tiller number under drought stress. Both QTNs resided in the intergenic region of the genome with a modifier effect on the traits.

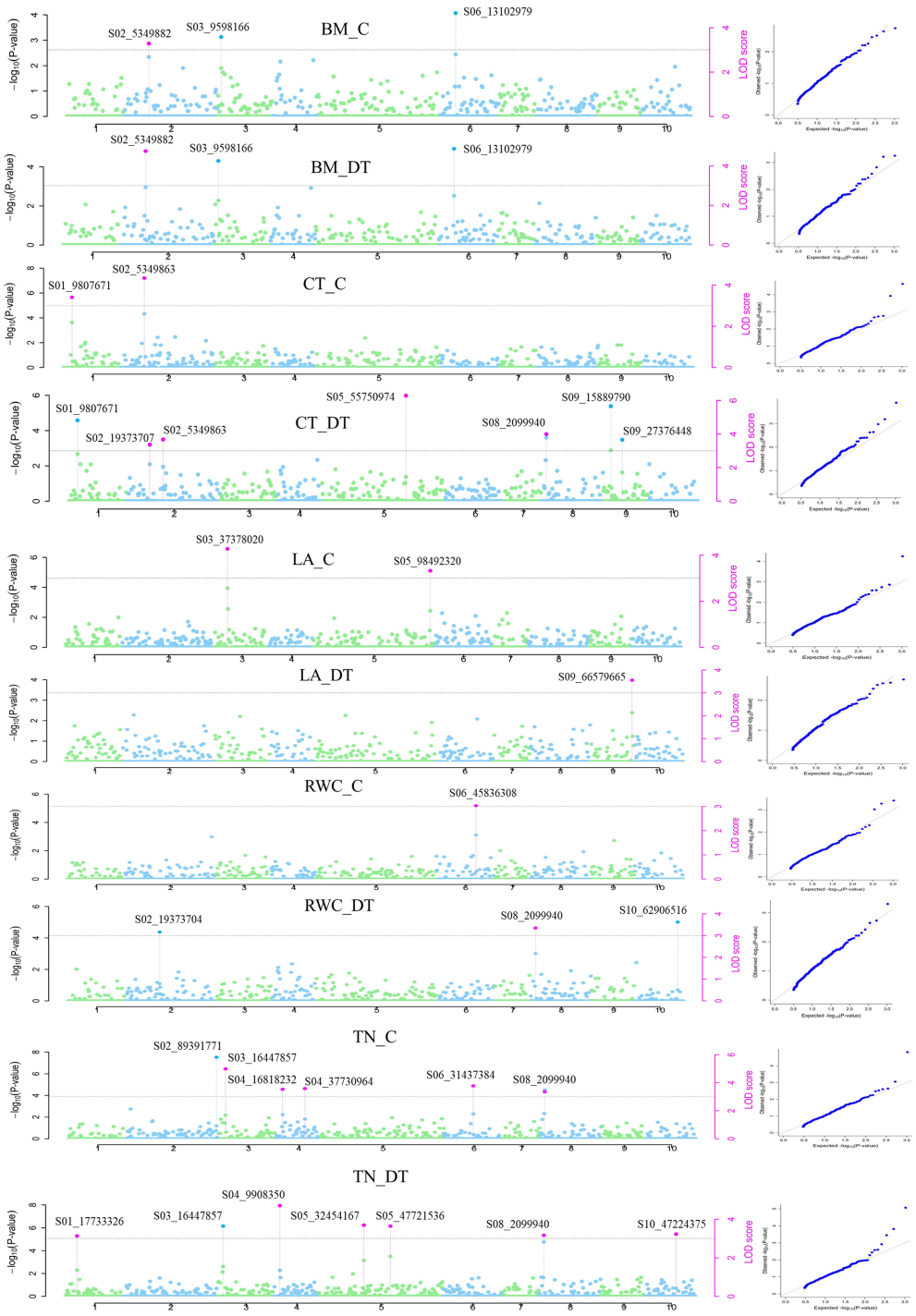


FIGURE 8
Manhattan plots showing significant QTN (purple color coded) with the corresponding QQ (quantile–quantile) plots representing the distribution of expected and observed *p*-values for various traits under control and drought conditions. CT_C and CT_ DT—Canopy temperature under control and drought conditions, respectively. LA_C and LA_DT—Leaf area under control and drought conditions, respectively. RWC_C and RWC_DT—Relative water content under control and drought conditions, respectively. TN_C and TN_DT—Tiller number under control and drought conditions, respectively. BM_C and BM_DT—Biomass under control and drought conditions, respectively.

Of the 34 unique genes from 235 genes identified from drought stress-responsive traits associated with 33 QTNs, 27 genes had known functional implications in drought perception and signaling. TOR 2, linked to the pleiotropic QTN (S08_2099940), belongs to the serine/threonine protein kinase family, which regulates signaling networks involved in cell growth and development and cellular metabolism (Bakshi et al., 2021). Overexpression of TOR was shown to increase the accumulation of osmolytes such as

proline to maintain cellular osmotic potential under abiotic stress conditions (Caldana et al., 2013). Candidate genes such as potassium transporter, auxin response factor, FAD-binding PCMH-type domain-containing protein, protein kinases, and EG45-like domain-containing protein were identified from QTNs associated with chlorophyll fluorescence under drought stress. Optimal potassium nutrition contributes to important biological functions to maintain osmoregulation and photosynthetic carbon metabolism when plants are exposed to drought stress conditions (Cakmak and Rengel, 2024). Genes encoding for potassium transporters and auxin response factors linked to QTNs associated with chlorophyll fluorescence in *Erianthus* clones might contribute to maintaining the photosynthetic efficiency and chlorophyll content under drought conditions (Verma et al., 2022). The receptor protein kinase TMK1 family protein identified in the study has been shown to be a key regulator of abscisic acid biosynthesis via auxin signaling under drought stress (Yang et al., 2021). NRT1/PTR family protein in the QTN associated with tiller number under drought stress is reported to increase nitrate reductase activity, contributing to increased root length and biomass production (Nedelyaeva et al., 2024).

A GWAS was performed in this study with 1,044 high-quality SNPs, which could be considered less given that the *Erianthus* clones used in the study had varying ploidy levels with 2n=2x=20 to 2n=2x=60. However, we believe that with the LD as large as ~3 Mb, the number of markers used were sufficient to capture the LD blocks. On the other hand, owing to budgetary constraints, the drought association panel investigated for GWAS was a small representative of the larger world collection of *Erianthus* clones. Although selective GBS of such a small-size population has been used to capture QTLs in sugarcane (Gutierrez et al., 2018), the use of the entire collection with more genotypic diversity will help capture better marker–trait associations.

Conclusion

Screening of *Erianthus* clones from a diverse collection of sugarcane and the related germplasm pool identified drought-tolerant clones with significant physiological adaptation traits that were able to maintain a higher photosynthetic efficiency to sustain growth and biomass accumulation under drought conditions. These clones can be used as important genetic resources in sugarcane breeding program to improve drought tolerance. The significant QTNs, especially the QTN (S08_2099940) associated with more than one trait, upon further validation using a larger-size and a more diverse population and/or interspecific cross progeny, can be used to develop diagnostic markers to facilitate marker-assisted transfer of drought tolerance-associated alleles into high-yielding but drought-sensitive sugarcane clones for the selection of drought-tolerant clones early in the variety development process. Similarly, genes such as target of rapamycin and serine/threonine protein kinase linked to the significant QTNs can be used to improve drought tolerance in sugarcane using gene overexpression and/or knock-out strategies.

Data availability statement

The sequence data analyzed in this study have been deposited to the NCBI SRA database and are available at https://www.ncbi.nlm.nih.gov/ (Bioproject PRJNA1368756).

Author contributions

VR: Conceptualization, Formal analysis, Funding acquisition, Investigation, Supervision, Writing – original draft. AP: Formal analysis, Methodology, Software, Visualization, Writing – original draft. PG: Formal analysis, Methodology, Writing – review & editing. AC: Formal analysis, Methodology, Writing – review & editing. HM: Formal analysis, Investigation, Methodology, Writing – review & editing. KM: Formal analysis, Methodology, Writing – review & editing. RR: Formal analysis, Methodology, Writing – review & editing. NB: Conceptualization, Funding acquisition, Investigation, Methodology, Supervision, Visualization, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research and/or publication of this article. This work was supported by the Department of Science and Technology, SERB Early Career Research Award (ECR/2016/001827/LS); the Department of Science and Technology, SERB SIRE fellowship (Grant number SIR/2022/001192); and the USDA-NIFA Hatch grant (Accession # 7006617).

Acknowledgments

The authors from the ICAR–Sugarcane Breeding Institute thank the Department of Biotechnology, Government of India, New Delhi, for supporting this research. RV is grateful to the Department of Science and Technology, SERB for supporting the visiting scientist position at the Louisiana State University Agricultural Center, USA through a SIRE fellowship. This manuscript has been approved for publication by the Louisiana Agricultural Experiment Station (MS#2025-306-40482) and ICAR Sugarcane Breeding Institute (contribution number 2025/-26/45).

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 handling editor GS declared a past co-authorship with the author AP.

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.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2025.1684712/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Percentage reduction in morphological and physiological traits among 223 *Erianthus* clones under drought stress compared to irrigated control condition.

SUPPLEMENTARY FIGURE 2
Percentage reduction in morphological and physiological traits among *Erianthus* drought panel under drought stress compared to irrigated control condition.

References

Augustine, S. M., Ashwin Narayan, J., Syamaladevi, D. P., Appunu, C., Chakravarthi, M., Ravichandran, V., et al. (2015). Overexpression of EaDREB2 and pyramiding of EaDREB2 with the pea DNA helicase gene (PDH45) enhance drought and salinity tolerance in sugarcane (*Saccharum* spp. hybrid). *Plant Cell Rep.* 34, 247–263. doi: 10.1007/s00299-014-1704-6

Bakshi, A., Moin, M., Madhav, M. S., Datla, R., and Kirti, P. B. (2021). Target of Rapamycin (TOR) negatively regulates chlorophyll degradation and lipid peroxidation and controls responses under abiotic stress in *Arabidopsis thaliana*. *Plant Stress* 2, 100020. doi: 10.1016/j.stress.2021.100020

Balsalobre, T. W. A., da Silva Pereira, G., Margarido, G. R. A., Gazaffi, R., Barreto, F. Z., Anoni, C. O., et al. (2017). GBS-based single dosage markers for linkage and QTL mapping allow gene mining for yield-related traits in sugarcane. *BMC Genomics* 18, 1–19. doi: 10.1186/s12864-016-3383-x

Bao, Y., Zhang, Q., Huang, J., Zhang, S., Yao, W., Yu, Z., et al. (2024). A chromosomal-scale genome assembly of modern cultivated hybrid sugarcane provides insights into origination and evolution. *Nat. Commun.* 15, 3041. doi: 10.1038/s41467-024-47390-6

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Society: Ser. B (Methodological)* 57, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x

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

Cai, Q., Aitken, K., Deng, H. H., Chen, X. W., Fu, C., and Jackson, P. A. (2005). Verification of the introgression of *Erianthus arundinaceus* germplasm into sugarcane using molecular markers. *Plant Breed.* 124, 322–328. doi: 10.1111/j.1439-0523.2005.01099.x

Cai, Q., Aitken, K. S., Fan, Y. H., Piperidis, G., Liu, X. L., McIntyre, C. L., et al. (2012). Assessment of the genetic diversity in a collection of *Erianthus arundinaceus*. *Genet. Resour. Crop Evol.* 59, 1483–1491. doi: 10.1007/s10722-011-9776-4

Cakmak, I., and Rengel, Z. (2024). Potassium may mitigate drought stress by increasing stem carbohydrates and their mobilization into grains. *J. Plant Physiol.* 303, 154325. doi: 10.1016/j.jplph.2024.154325

Caldana, C., Li, Y., Leisse, A., Zhang, Y., Bartholomaeus, L., Fernie, A. R., et al. (2013). Systemic analysis of inducible target of rapamycin mutants reveals a general metabolic switch controlling growth in *Arabidopsis thaliana*. *Plant J.* 73, 897–909. doi: 10.1111/tpj.12080

Chakraborty, D., Garg, R. N., Tomar, R. K., Dwivedi, B. S., Aggarwal, P., Singh, R., et al. (2010). Soil physical quality as influenced by long-term application of fertilizers and manure under maize-wheat system. *Soil Sci.* 175, 128–136. doi: 10.1097/SS.0b013e3181d53bd7

Chang, D., Yang, F. Y., Yan, J. J., Wu, Y. Q., Bai, S. Q., Liang, X. Z., et al. (2012). SRAP analysis of genetic diversity of nine native populations of wild sugarcane, *Saccharum spontaneum*, from Sichuan, China. *Genet. Mol. Res.* 11, 1245–1253. doi: 10.4238/2012

Cortes, J. D., Gutierrez, A. F., Hoy, J. W., Hale, A. L., and Baisakh, N. (2024). Genetic mapping of quantitative trait loci controlling smut resistance in Louisiana sugarcane using a bi-parental mapping population. *Plant Gene* 37, 100445. doi: 10.1016/j.plgene.2023.100445

D'Hont, A., Rao, P. S., Feldmann, P., Grivet, L., Islam-Faridi, N., and Taylor, P. (1995). Identification and characterisation of sugarcane intergeneric hybrids, *Saccharum officinarum* x *Erianthus arundinaceus*, with molecular markers and DNA in *situ* hybridisation. *Theor. Appl. Genet.* 91, 320–326. doi: 10.1007/BF00220894

Dao, Z. X., Yan, G. J., Zhang, J. B., Chang, D., Bai, S. J., Chen, Z. H., et al. (2013). Investigation and collection of wild *Erianthus arundinaceum* germplasm resources. *J. Plant Genet. Resour.* 14, 816–820. doi: 10.13430/j.cnki.jpgr.2013.05.008

Deng, H. H., Liao, Z. Z., Li, Q. W., Lao, F. Y., Fu, C., and Chen, X. W. (2002). Breeding and isozyme marker assisted selection of F2 hybrids from Saccharum spp. x *Erianthus arundinaceus*. *Sugarcane Canesugar.* 1, 1–5.

Dhansu, P., Kulshreshtha, N., Kumar, R., Raja, A. K., Pandey, S. K., Goel, V., et al. (2021). Identification of drought-tolerant co-canes based on physiological traits, yield attributes and drought tolerance indices. *Sugar Tech* 23, 747–761. doi: 10.1007/s12355-021-00932-0

Fernandez, G. C. J. (1992). "Effective selection criteria for assessing stress tolerance," in *Proceedings of the international symposium on adaptation of vegetables and other food crops to temperature and water stress*. Ed. C. G. Kuo (Tainan, Taiwan: Asian Vegetable Research and Development Center), 257–270.

Ferreira, T. H., Tsunada, M. S., Bassi, D., Araújo, P., Mattiello, L., and Guidelli, G. V. (2017). Sugarcane water stress tolerance mechanisms and its implications on developing biotechnology solutions. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.01077

Fickett, N., Gutierrez, A., Verma, M., Pontif, M., Hale, A., Kimbeng, C., et al. (2019). Genome-wide association mapping identifies markers associated with cane yield components and sucrose traits in the Louisiana sugarcane core collection. *Genomics* 111, 1794–1801. doi: 10.1016/j.ygeno.2018.12.002

Fischer, R. A., and Maurer, R. (1978). Drought resistance in spring wheat cultivars. I. Grain yield responses. *Aust. J. Agric. Res.* 29, 897–912. doi: 10.1071/AR9780897

Fukuhara, S., Terajima, Y., Irei, S., Sakaigaichi, T., Ujihara, K., Sugimoto, A., et al. (2013). Identification and characterization of intergeneric hybrid of commercial sugarcane (*Saccharum* spp. hybrid) and *Erianthus arundinaceus* (Retz.) Jeswiet. *Euphytica* 189, 321–327. doi: 10.1007/s10681-012-0748-3

Garsmeur, O., Droc, G., Antonise, R., Grimwood, J., Potier, B., Aitken, K., et al. (2018). A mosaic monoploid reference sequence for the highly complex genome of sugarcane. *Nat. Commun.* 9, 2638. doi: 10.1038/s41467-018-05051-5

Gutierrez, A. F., Hoy, J. W., Kimbeng, C. A., and Baisakh, N. (2018). Identification of genomic regions controlling leaf scald resistance in sugarcane using a bi-parental mapping population and selective genotyping by sequencing. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.00877

Guttieri, M. J., Stark, J. C., O'Brien, K., and Souza, E. (2001). Relative sensitivity of spring wheat grain yield and quality parameters to moisture deficit. *Crop Sci.* 41, 327–335. doi: 10.2135/cropsci2001.412327x

Healey, A. L., Garsmeur, O., Lovell, J. T., Shengquiang, S., Sreedasyam, A., Jenkins, J., et al. (2024). The complex polyploid genome architecture of sugarcane. *Nature* 628, 804–810. doi: 10.1038/s41586-024-07231-4

Hemaprabha, G., Swapna, S., Lavanya, D. L., and Srinivasan, T. V. (2013). Evaluation of drought tolerance potential of elite genotypes and progenies of sugarcane (*Saccharum* sp. hybrids). *Sugar Tech* 15, 9–16. doi: 10.1007/s12355-012-0182-9

Hoang, D. T., Hiroo, T., and Yoshinobu, K. (2019). Nitrogen use efficiency and drought tolerant ability of various sugarcane varieties under drought stress at early growth stage. *Plant Production Sci.* 22, 250–261. doi: 10.1080/1343943X.2018.1540277

Jackson, P., and Henry, R. J. (2011). "Erianthus," in *Wild crop relatives: Genomic and breeding resources: Industrial crops* (Springer Berlin Heidelberg, Berlin, Heidelberg), 97–107. doi: 10.1007/978-3-642-21102-7_5

Jombart, T., Devillard, S., and Balloux, F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet.* 11, 1–15. doi: 10.1186/1471-2156-11-94

Kumar, S., Wang, S., Wang, M., Zeb, S., Khan, M. N., Chen, Y., et al. (2024). Enhancement of sweetpotato tolerance to chromium stress through melatonin and glutathione: Insights into photosynthetic efficiency, oxidative defense, and growth parameters. *Plant Physiol. Biochem.* 208, 108509. doi: 10.1016/j.plaphy.2024.108509

Lan, J. (1998). Comparison of evaluating methods for agronomic drought resistance in crops. *Acta Agriculturae Boreali-occidentalis Sin.* 7, 85–87.

Meena, M. R., Govindaraj, P., Kumar, R. A., Elayaraja, K., Appunu, C., Kumar, R., et al. (2024). Biomass and energy potential of *Erianthus arundinaceus* and *Saccharum spontaneum*-derived novel sugarcane hybrids in rainfed environments. *BMC Plant Biol.* 24, 198. doi: 10.1186/s12870-024-04885-0

Mohanraj, K., Hemaprabha, G., and Vasantha, S. (2021). Biomass yield, dry matter partitioning and physiology of commercial and *Erianthus* introgressed sugarcane clones under contrasting water regimes. *Agric. Water Manage.* 255, 107035. doi: 10.1016/j.agwat.2021.107035

Montgomery, E. G. (1911). "Correlation studies in corn," in *Nebraska agricultural experiment station annual report*, Lincoln, Nebraska 24, 108–159.

Mukherjee, S. K. (1957). Origin and distribution of saccharum. *Botanical Gazette* 119, 55–61. doi: 10.1086/335962

Mukherjee, S. K. (1958). Revision of the genus *erianthus*. *Lloydia* 21, 157–188.

Nair, N. V., and Mary, S. (2006). RAPD analysis reveals the presence of mainland Indian and Indonesian forms of *Erianthus arundinaceus* (Retz.) Jeswiet in the Andaman–Nicobar Islands. *Curr. Sci.* 90, 1118–1122.

Nair, N. V., Mohanraj, K., Sunadaravelpandian, K., Suganya, A., Selvi, A., and Appunu, C. (2017). Characterization of an intergeneric hybrid of *Erianthus procerus*× *Saccharum officinarum* and its backcross progenies. *Euphytica* 213, 1–11. doi: 10.1007/s10681-017-2053-7

Narayanan, S., Mohan, A., Gill, K. S., and Prasad, P. V. (2014). Variability of root traits in spring wheat germplasm. *PloS One* 9, e100317. doi: 10.1371/journal.pone.0100317

Nedelyaeva, O. I., Khramov, D. E., Balnokin, Y. V., and Volkov, V. S. (2024). Functional and molecular characterization of plant nitrate transporters belonging to NPF (NRT1/PTR) 6 subfamily. *Int. J. Mol. Sci.* 25, 13648. doi: 10.3390/ijms252413648

Pachakkil, B., Terajima, Y., Ohmido, N., Ebina, M., Irei, S., Hayashi, H., et al. (2019). Cytogenetic and agronomic characterization of intergeneric hybrids between Saccharum spp. hybrid and *Erianthus arundinaceus*. *Sci. Rep.* 9, 1748. doi: 10.1038/s41598-018-38316-6

Pierre, J. S., Perroux, J. M., and Rae, A. L. (2019). Screening for sugarcane root phenes reveals that reducing tillering does not lead to an increased root mass fraction. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.00119

Piperidis, G., Christopher, M. J., Carroll, B. J., Berding, N., and D'Hont, A. (2000). Molecular contribution to selection of intergeneric hybrids between sugarcane and the wild species *Erianthus arundinaceus*. *Genome* 43, 1033–1037. doi: 10.1139/g00-059

Piperidis, G., and D'hont, A. (2001). *Chromosome composition analysis of various Saccharum interspecific hybrids by genomic in* situ *hybridisation (GISH)* Vol. 2 (Brisbane, Australia: Australian Society of Sugar Cane Technologists), 565–566.

Piperidis, G., Piperidis, N., and D'Hont, A. (2010). Molecular cytogenetic investigation of chromosome composition and transmission in sugarcane. *Mol. Genet. Genomics* 284, 65–73. doi: 10.1007/s00438-010-0546-3

Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959. doi: 10.1093/genetics/155.2.945

Ram, B., Sreenivasan, T. V., Sahi, B. K., and Singh, N. (2001). Introgression of low temperature tolerance and red rot resistance from *Erianthus* in sugarcane. *Euphytica* 122, 145–153. doi: 10.1023/A:1012626805467

Reyes, J. A. O., Casas, D. E., Gandia, J. L., and Delfin, E. F. (2021). Drought impact on sugarcane production. *Agric. Res. Updates* 35, 53–93.

Reynolds, M. P., Hays, D., and Chapman, S. (2010). "Breeding for adaptation to heat and drought stress," in *Climate change and crop production*. Ed. M. P. Reynolds (Oxfordshire, UK: CABI), 71–91.

Sanghera, G. S., and Kashyap, L. (2022). Stress indices based on cane and sugar yields: Implications in selection of sugarcane varieties for drought tolerance. *Agric. Res. J.* 59, 816–834. doi: 10.5958/2395-146X.2022.00116.8

Scortecci, K. C., Creste, S., Calsa, T. Jr., Xavier, M. A., Landell, M. G., Figueira, A., et al. (2012). Challenges, opportunities and recent advances in sugarcane breeding. *Plant Breed.* 1, 267–296.

Siddique, M. R. B., Hamid, A. I. M. S., and Islam, M. S. (2000). Drought stress effects on water relations of wheat. *Botanical Bull. Academia Sin.* 41, 35–39.

Silva, M. D. A., Jifon, J. L., Da Silva, J. A., and Sharma, V. (2007). Use of physiological parameters as fast tools to screen for drought tolerance in sugarcane. *Braz. J. Plant Physiol.* 19, 193–201. doi: 10.1590/S1677-04202007000300003

Sreenivasan, T. V., Ahloowalia, B. S., and Heinz, D. J. (1987). "Cytogenetics," in *Sugarcane improvement through breeding*. Ed. D. J. Heinz (Amsterdam, Netherlands: Elsevier).

Takaragawa, H., and Wakayama, M. (2024). Responses of leaf gas exchange and metabolites to drought stress in different organs of sugarcane and its closely related species *Erianthus arundinaceus*. *Planta* 260, 90. doi: 10.1007/s00425-024-04536-6

Terajima, Y., Ponragdee, W., Sansayawichai, T., Tippayawat, A., Chanachai, S., Ebina, M., et al. (2022). Genetic variation in agronomic traits of *Erianthus* germplasm under multiple-ratoon crops in Thailand. *Crop Sci.* 62, 1531–1549. doi: 10.1002/csc2.20697

Terajima, Y., Sugimoto, A., Tippayawat, A., Irei, S., and Hayashi, H. (2023). Root distribution and fibre composition of intergeneric F1 hybrid between sugarcane and *E. arundinaceus*. *Field Crops Res.* 197, 108920. doi: 10.1016/j.fcr.2023.108920

Tsuruta, S., Masumi, E., Makoto, K., Taiichiro, H., and Takayoshi, T. (2012). Analysis of genetic diversity in the bioenergy plant *Erianthus arundinaceus* (Poaceae: Andropogoneae) using amplified fragment length polymorphism markers. *Grassland Sci.* 58, 174–177. doi: 10.1111/j.1744-697X.2012.00258.x

Tsuruta, S., Srithawong, S., Sakunrungsirikul, S., Ebina, M., Kobayashi, M., et al. (2022). *Erianthus* germplasm collection in Thailand: genetic structure and phylogenetic aspects of tetraploid and hexaploid accessions. *BMC Plant Biol.* 22, 45. doi: 10.1186/s12870-021-03418-3

Valarmathi, R., Mahadeva Swamy, H. K., Appunu, C., Suresha, G. S., Mohanraj, K., Hemaprabha, G., et al. (2023). Comparative transcriptome profiling to unravel the key molecular signalling pathways and drought adaptive plasticity in shoot borne root system of sugarcane. *Sci. Rep.* 13, 12853. doi: 10.1038/s41598-023-39970-1

Valarmathi, R., MahadevaSwamy, H. K., Preethi, K., Narayan, J. A., Appunu, C., and Rahman, H. (2020). Characterization and in silico analyses of RTCS gene from sugarcane encoding LOB protein family of transcription factors: a key regulator of shoot-borne root initiation. *J. Sugarcane Res.* 10, 12–23. doi: 10.37580/JSR.2020.1.10.12-23

Valarmathi, R., Mahadevaswamy, H. K., Ulaganathan, V., Appunu, C., Hemaprabha, G., et al. (2021). Low cost high throughput image based root phenotyping pipeline for evaluation of sugarcane root system architecture under drought stress. *J. Sugarcane Res.* 11, 24–36. doi: 10.37580/JSR.2021.1.11.24-36

Verma, K. K., Song, X. P., Budeguer, F., Nikpay, A., Enrique, R., Singh, M., et al. (2022). Genetic engineering: An efficient approach to mitigating biotic and abiotic stresses in sugarcane cultivation. *Plant Signaling Behav.* 17, 2108253. doi: 10.1080/15592324.2022.2108253

Vinoth, P., Ariharasutharsan, G., Malarvizhi, A., Senthilrajan, P., Appunu, C., Vinu, V., et al. (2022). Comparative leaf, root morpho-anatomical phenes in sugarcane wild germplasm and commercial genotypes conferring drought and salinity stress tolerance. *J. Sugarcane Res.* 12, 146–161. doi: 10.37580/JSR.2022.2.12.146-161

Wang, W., Li, R., Wang, H., Qi, B., Jiang, X., Zhu, Q., et al. (2019). Sweetcane (*Erianthus arundinaceus*) as a native bioenergy crop with environmental remediation potential in southern China: A review. *GCB Bioenergy* 11, 1012–1025. doi: 10.1111/gcbb.12600

Wang, X., Liu, H., Ma, H., Dang, Y., Han, Y., Zhang, C., et al. (2025). Photosynthetic performance and sucrose metabolism in superior and inferior rice grains with overlapping growth stages under water stress. *Sci. Rep.* 15, 11973. doi: 10.1371/journal.pone.0324128

Wang, W., Vinocur, B., and Altman, A. (2003). Plant responses to drought, salinity and extreme temperatures: towards genetic engineering for stress tolerance. *Planta* 218, 1–14. doi: 10.1007/s00425-003-1105-5

Wang, T., Wang, B., Hua, X., Tang, H., Zhang, Z., Gao, R., et al. (2023). A complete gap-free diploid genome in *Saccharum* complex and the genomic footprints of evolution in the highly polyploid *Saccharum* genus. *Nat. Plants* 9, 554–571. doi: 10.1038/s41477-023-01378-0

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

Wirojsirasak, W., Songsri, P., Jongrungklang, N., Tangphatsornruang, S., Klomsa-Ard, P., and Ukoskit, K. (2023). A large-scale candidate-gene association mapping for drought tolerance and agronomic traits in sugarcane. *Int. J. Mol. Sci.* 24, 12801. doi: 10.3390/ijms241612801

Yang, J., He, H., He, Y., Zheng, Q., Li, Q., Feng, X., et al. (2021). TMK1-based auxin signaling regulates abscisic acid responses via phosphorylating ABI1/2 in Arabidopsis. *Proc. Natl. Acad. Sci. U. S. A* 118, e2102544118. doi: 10.1073/pnas.2102544118

Yang, X., Islam, M. S., Sood, S., Maya, S., Hanson, E. A., Comstock, J., et al. (2018). Identifying quantitative trait loci (QTLs) and developing diagnostic markers linked to orange rust resistance in sugarcane (*Saccharum* spp.). *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.00350

Yang, X., Luo, Z., Todd, J., Sood, S., and Wang, J. (2020). Genome-wide association study of multiple yield traits in a diversity panel of polyploid sugarcane (Saccharum spp.). *Plant Genome* 13, e20006. doi: 10.1002/tpg2.20006

Yang, X., Todd, J., Arundale, R., Binder, J. B., Luo, Z., Islam, M. S., et al. (2019). Identifying loci controlling fiber composition in polyploid sugarcane (*Saccharum* spp.) through genome-wide association study. *Ind. Crops Products* 130, 598–605. doi: 10.1016/j.indcrop.2019.01.023

Yao, J., Sun, D., Cen, H., Xu, H., Weng, H., Yuan, F., et al. (2018). Phenotyping of Arabidopsis drought stress response using kinetic chlorophyll fluorescence and multicolor fluorescence imaging. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.00603

Zhang, J., Zhang, X., Tang, H., Zhang, Q., Hua, X., Ma, X., et al. (2018). Allele-defined genome of the autopolyploid sugarcane *Saccharum spontaneum* L. *Nat. Genet.* 50, 1565–1573. doi: 10.1038/s41588-018-0237-2

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 09 December 2025
DOI 10.3389/fpls.2025.1687331

OPEN ACCESS

EDITED BY
Santosh Gudi,
North Dakota State University, United States

REVIEWED BY
Krishna Sai Karnatam,
West Virginia State University, United States
Rakshith S. R. Gowda,
Murdoch University, Australia
Mankanwal Goraya,
The Pennsylvania State University,
United States

*CORRESPONDENCE
Satish Kumar
kumarsatish227@gmail.com

RECEIVED 17 August 2025
ACCEPTED 24 October 2025
PUBLISHED 09 December 2025
CORRECTED 11 December 2025

CITATION
Tanwar V, Kumar S, Lal C, Aggarwal R, Nair R, Kamboj D, Kashyap PL, Singh V, Saini JS, Kashyap S, Wani SH, Udupa SM, Singh R and Tiwari R (2025) Genome-wide association studies for identification of stripe rust resistance loci in diverse wheat genotypes. *Front. Plant Sci.* 16:1687331. doi: 10.3389/fpls.2025.1687331

Genome-wide association studies for identification of stripe rust resistance loci in diverse wheat genotypes

Vikesh Tanwar[1], Satish Kumar[1]*, Chuni Lal[1], Rajesh Aggarwal[1], Rajitha Nair[1], Disha Kamboj[1], Prem Lal Kashyap[1], Vikram Singh[2], Johar Singh Saini[3], Sunil Kashyap[3], Shabir Hussain Wani[4], Sripada M. Udupa[5], Rajender Singh[1] and Ratan Tiwari[1]

[1]Indian Council of Agricultural Research (ICAR)-Indian Institute of Wheat and Barley Research, Karnal, India, [2]Chaudhary Charan Singh (CCS) Haryana Agricultural University, Hisar, India, [3]Punjab Agricultural University, Regional Research Station, Gurdaspur, India, [4]Sher-e-Kashmir University of Agricultural Sciences and Technology-Kashmir (SKUAST-K), Srinaga, India, [5]International Centre for Agricultural Research in the Dry Areas, Rabat, Morocco

Introduction: In North India, *Puccinia striiformis* f. sp. *triticii* (Pst), the causal agent of stripe rust, poses a significant challenge to wheat productivity. The frequent emergence of new virulent *Pst* strains has rendered many resistance genes ineffective. Hence, continuous identification and deployment of novel resistance genes are crucial for maintaining durable resistance and ensuring sustainable wheat cultivation.

Materials and Methods: A genome-wide association study (GWAS) was conducted on 652 elite, diverse wheat genotypes using 1,938 DArTseq SNP markers. Field evaluations were performed at the adult plant stage across four locations—Hisar, Karnal, Gurdaspur, and Khudwani—under natural disease conditions. Marker–trait associations were identified using General Linear Model (GLM), Mixed Linear Model (MLM), and FarmCPU approaches, considering loci with $-\log_{10}(p) \geq 3$ as significant.

Results: Analysis revealed 27 genomic regions significantly associated with stripe rust resistance across environments. Among these, four loci were located on chromosomes 2B and 6B, and three on 6A. Several loci corresponded to resistance-related genes, including NBS-LRR, F-box, LRR, protein kinase, Ser/Thr_kinase, Znf_RING-CH, E3-ubiquitin ligase, and ABC transporter genes, suggesting their potential role in rust resistance mechanisms.

Discussion: The study identified novel genomic regions associated with Pst resistance, providing valuable resources for wheat improvement. The functional annotation of these loci highlights their involvement in plant defense pathways. Conversion of these loci into breeder-friendly molecular markers will facilitate marker-assisted selection (MAS) and accelerate the development of durable stripe rust-resistant wheat cultivars suited to North Indian agro-ecological conditions.

KEYWORDS

wheat, biotic stress, stripe rust, screening, resistance

Introduction

Wheat (*Triticum aestivum* L.) is a key staple crop that plays an essential role in feeding the global population and supporting food security. It is cultivated worldwide over an area of approximately 220 million hectares, yielding approximately 775.4 million tonnes of annual production (Prospects, 2022). By 2050, the global population is expected to reach approximately 9 billion, leading to a significant increase in demand for wheat. This crop is a vital source of daily caloric and protein intake, accounting for over 20% of caloric consumption and 25% of the protein on a global scale (Hubert et al., 2010). Nevertheless, wheat cultivation faces significant abiotic and biotic challenges that result in considerable reductions in yield and quality (Trethowan et al., 2018). According to Savary et al. (2019), it is estimated that diseases alone could lead to yield losses of approximately 21.5%. Among major wheat diseases, rusts are the most prevalent and damaging threats to wheat crops, affecting all regions where wheat is cultivated.

Among the various types of rust, stripe rust, commonly known as yellow rust, is a major threat to bread wheat (*T. aestivum* L.) on a global scale, caused by the fungus *Puccinia striiformis* f. sp. *tritici* (*Pst*). This disease significantly impacts wheat production worldwide. An outbreak of this disease rapidly devastates green leaves, leading to a drastic reduction in photosynthesis. Consequently, plants become weakened and stunted, resulting in fewer grains per spike, shriveled grains, and lower grain weights. According to Chen (2020), fields planted with susceptible cultivars are at high risk, with potential grain yield losses reaching up to 100%. Currently, 5.47 million tons of annual global losses in grain yield occur due to this disease (Beddow et al., 2015). Some fungicides effectively control this disease, but their use typically leads to higher crop production costs and poses significant environmental risks. Therefore, harnessing genetic resistance is the most powerful, cost-effective, and environmentally sustainable strategy for effectively combating this disease (Mapuranga et al., 2022). It is essential for researchers to actively seek out new resistance sources and incorporate innovative resistance genes into cultivars to stay ahead of emerging races of stripe rust. This proactive strategy will not only control the disease but also eliminate the "boom and bust cycle" associated with the performance of cultivars.

According to Van der Plank (Parlevliet and Zadoks, 1977), plant resistance is classified into two primary types—horizontal and vertical—which are determined by the interaction between the pathogen and host. Horizontal resistance, often referred to as partial resistance or non-specific resistance, is characterized by its ability to provide intermediate protection against a broad range of pathogens. This type of resistance is regulated by multiple genes and is fundamentally quantitative, resulting in various levels of resistance rather than absolute immunity. Furthermore, several biochemical and physiological processes, including the production of antimicrobial compounds and the reinforcement of cell walls, significantly contribute to the effective inhibition of pathogen colonization (Dyakov, 2007). Vertical resistance, also referred to as complete or specific resistance, represents a defense mechanism inherent in plants, providing robust protection against specific strains of pathogens or a limited range of closely related strains. This form of resistance is primarily regulated by one or a few genes, offering the plant comprehensive immunity to the particular pathogen in question. Specific resistance (R) genes in the plant's genome enable the synthesis of proteins that recognize and interact directly with pathogen compounds, resulting in a prompt and accurate defensive reaction. This defensive response may encompass the activation of defense-related genes, the synthesis of toxic compounds, or the induction of localized cell death at the infection site, all of which serve to inhibit the pathogen's proliferation (Orton, 2020). Chen (2005) classified host resistance into two categories: adult-plant resistance (APR), which is defined as horizontal-type resistance and regulated by numerous genes, and seedling or all-stage resistance (ASR), which is controlled by one or a few genes, similar to vertical-type resistance. ASR is effective during all growth stages and is typically characterized as having qualitative or monogenic resistance. ASR adheres to the gene-for-gene model described by Flor (1971), offering high levels of protection, but it is race-specific, and its effectiveness is compromised by high selection pressure on the pathogen, which may mutate to overcome resistance. In contrast, APR is more durable but often provides partial resistance. It is expressed or enhanced at the adult plant stage (Lagudah, 2011; Chen, 2013; Mundt, 2014; Ellis et al., 2014). Moreover, APR is usually non-race specific, although race-specific APR has also been identified (Milus et al., 2015). Despite their durability, APR genes do not protect plants at the seedling stage and tend to show variability in the timing and levels of resistance across environments, where a single APR gene often provides insufficient protection under severe epidemics (Risk et al., 2012; Chen, 2014; Singh et al., 2015). Pyramiding multiple APR genes is essential to provide a high level of resistance through additive or epistatic effects (Sørensen et al., 2014). Plant breeders seek to incorporate both forms of resistance into cultivars through systematic breeding programs to improve disease resistance in crops. In the agricultural sector, this approach is fundamental to integrated and sustainable pest management techniques (Sharma et al., 2025).

Eighty-seven genes are resistant to stripe rust (*Yr1* to *Yr87*), and over 350 quantitative trait loci (QTLs) have been identified and mapped to the wheat genome thus far (McIntosh et al., 2020; Zhu et al., 2023; Sharma et al., 2024). Out of the total named stripe rust resistance genes, 59 provide resistance at the seedling stage, while 28 provide resistance at the adult plant stage (APR). To date, only 12 *Yr* genes have been cloned, namely, *Yr5*, *Yr7*, *Yr10* (*YrNAM*), *Yr15*, *Yr18*, *Yr27*, *Yr36*, *Yr46*, *YrSP*, *YrAS2388*, *YrU1*, and *Yr87* (Fu et al., 2009; Krattinger et al., 2009; Liu et al., 2014; Moore et al., 2015; Marchal et al., 2018; Klymiuk et al., 2018; Zhang et al., 2019; Wang et al., 2020; Athiyannan et al., 2022; Ni et al., 2023; Sharma et al., 2024), of which *Yr18*, *Yr36*, and *Yr46* are APR genes. Generally, ASR genes are associated with nucleotide-binding domain and leucine-rich repeat proteins (Sánchez-Martín and Keller, 2021). These proteins recognize effector proteins produced by the pathogen to initiate effector-triggered immunity, thereby protecting the host (Gururani et al., 2012). In contrast, APR

genes lack specific structural domains, feature more complex structures, and indirectly contribute to resistance (Sánchez-Martín and Keller, 2021). For example, *Yr18* encodes an ATP-binding cassette (ABC) transporter, *Yr36* encodes a protein kinase (WKS1), and *Yr46* encodes a hexose transporter. Most ASR genes and some APR genes deployed in commercial wheat cultivars are no longer effective due to the emergence of virulent *Pst* races (Hovmøller et al., 2011; Sørensen et al., 2014; Wan and Chen, 2014). Therefore, finding user-friendly markers becomes even more necessary (Zeng et al., 2019).

With the sequencing technology advancements in recent years, this has facilitated the development of innovative genotyping approaches that provide cost-effective and high-throughput capabilities for marker systems. Among these methods, Diversity Array Technology Single Nucleotide Polymorphism (DArTSNP) markers are particularly well-suited for diversity analysis, marker-assisted genetic resource management, marker-assisted selection, and genome-wide association studies (GWASs) in breeding programs. Through GWASs, progress in the identification of genes associated with stripe rust has been made in recent years. This is a robust methodology for elucidating the genetic foundation of complex traits. By analyzing extensive genetic data from diverse populations, researchers have identified specific genomic regions associated with resistance to stripe rust. These studies led to the discovery of novel genes and alleles of resistance that contribute to various agronomic traits, including drought and heat tolerance (Tadesse et al., 2019; Ballesta et al., 2019; Gudi et al., 2024, Gudi et al., 2025) as well as disease resistance (Kankwatsa et al., 2017; Kang et al., 2020) and grain quality characteristics (Tadesse et al., 2015). Many other complex agronomic traits in bread wheat provide valuable insights into understanding the plant immunity mechanisms. Finding these genes is a crucial step in creating improved and better cultivars with strong resistance to stripe rust, which will result in far more productive and sustainable agricultural practices. According to Brachi et al. (2011), GWASs effectively overcome the common challenges associated with bi-parental QTL mapping, such as restricted allelic diversity and limited genomic regions. The genetic architecture of complex traits in diverse germplasm collections can be studied using GWASs, which detect the genomic regions present in linkage disequilibrium (LD) with genes associated with the trait under study (Hall et al., 2010; Zhao et al., 2011; Riedelsheimer et al., 2012; Singh et al., 2025). Also, GWASs have been utilized to detect stripe rust resistance loci in different market classes of wheat (Naruoka et al., 2015; Liu et al., 2018; Liu et al., 2020; Mu et al., 2020; Muleta et al., 2020; Aoun et al., 2021; Zhang et al., 2021; Jambuthenne et al., 2022; El-Messoadi et al., 2024; Gao et al., 2024; Qiao et al., 2024). This study, consisting of 652 genotypes, also focused on identifying diverse wheat genotypes that revealed strong resistance at the adult plant stage to stripe rust, as well as finding genomic regions associated with stripe rust resistance through GWASs. A large and diverse set of wheat germplasm, including pre-bred materials, registered genetic stocks, and landraces, were used to dissect the variation for stripe rust resistance across four hotspot locations. A mid-density Single Nucleotide Polymorphism (SNP) marker panel was also used to scan the entire genome, which will reveal potentially novel QTLs that are different from known resistance genes. These identified QTLs can be further used for marker-assisted selection to develop improved, resistant wheat cultivars.

Materials and methods

Plant materials

The present study comprised a diverse set of 652 wheat genotypes. These genotypes included varieties (389), advanced breeding lines (53), genetic stocks (80), landraces (103), exotic lines (1), and mutants (24) (Supplementary Table 1). Genotypes in the panel are adapted to different agro-climatic zones of India, and this panel includes genotypes from all over the world. The stripe rust disease was evaluated in the field at four locations, namely, Karnal, Hisar, Gurdaspur, and Khudwani, in India during 2022–2023 (Table 1). These locations are in the stripe rust-prone areas of Northern India, which are suitable for studying natural disease pressure. These genotypes were obtained from the Germplasm Resource Unit (GRU), ICAR-IIWBR, Karnal, India.

Stripe rust evaluation at the adult plant stage in the field

The genotypes were sown in an augmented block design with 1-m length of a single row and 0.4-m distance between the rows. Sowing was performed in the first fortnight of November 2022–2023 at all four locations mentioned above. A variety of check lines that are known to be vulnerable to different rust infections were planted across the plot in infector rows (every 20th single row) and spreader rows (perpendicular to the 1-m rows). This approach establishes a robust inoculum, ensuring uniform disease development throughout the area. To achieve optimal disease distribution, collection was performed to capture naturally occurring disease during early infection in the spreader rows and used to inoculate the infector rows. The response to rust disease infection was measured using two parameters, namely, disease severity (DS) and infection response (IR). DS was assessed using the modified Cobb's scale (Peterson et al., 1948), which estimated the percentage coverage of rust pustules (uredinia) on the flag leaf.

TABLE 1 Field and geographical location of all the locations where experiment was performed.

Field location	Geographical location
New Farm, ICAR-IIWBR, Karnal, India	29°42′10.0″N, 76°59 29.7″E
Wheat and Barley Section, CCSHAU, Hisar, India	29°8 22.1928″ N, 75°4253.8812 E
RRS, PAU Gurdaspur, India	30°03′N, 75°27′E
RRS, SKAUST, Khudwani, India	34°17′N, 75°27′E

The scale ranged from 0 to 100, with 0 indicating no coverage and 100 indicating full coverage. IR was evaluated by measuring the host's reaction to rust pustules and converting it into a 0–1 scale (Roelfs et al., 1992). A fifth group, not mentioned by Roelfs et al. (1992), was classified for lines with a mixed response, ranging from moderately resistant to moderately susceptible. Rust response was evaluated using five scoring categories: resistant (R) at 0.2, moderately resistant (MR) at 0.4, mixed response (M) at 0.6, moderately susceptible (MS) at 0.8, and susceptible (S) at 1. Data were collected weekly, and three observations were made when the disease score of 60S (DS: 60; IR: S) was reached by the flag leaves of susceptible checks. Out of these multiple scores of a test line, the one with the score tending toward susceptibility from the last week was kept for the study. These scores were then utilized for GWASs across each environment by consolidating the two metrics into a single value known as the coefficient of infection (CI). This value was calculated as the product of DS and infection rate (IR) on a linear scale of 0 to 100, following the established methodologies of Loegering (1959) and Roelfs et al. (1992). Genotypes with CI scores of 0 to 1 were considered immune, 1–5 highly resistant, 5–15 moderately resistant, 15–25 moderately susceptible, and >25 susceptible. As a key trait for detecting significant marker–trait associations (MTAs), CI is very well-suited for GWASs and efficiently integrates data from both DS and IR about rust response (Yu et al., 2012; Gao et al., 2017; Mihalyov et al., 2017).

Statistical analysis

ANOVA

An augmented RCBD package was used to analyze variance [analysis of variance (ANOVA)] for 652 genotypes across all locations using the R software (*https://www.r-project.org*; Aravind et al., 2020).

DNA isolation and genotyping

At the growth stage (GS), 12 of the plant, genomic DNA was isolated (Zadoks et al., 1974), and the modified Cetyl Trimethlyl Ammonium Bromide (CTAB) method was applied (Saghai-Maroof et al., 1984) for the isolation of DNA. Samples were subjected to snap freezing and grinding using liquid nitrogen. Following RNase treatment, aqueous phase separation was conducted at a centrifugal force of 13,000 rpm for 10 minutes at 4°C, utilizing a phenol: chloroform:isoamyl alcohol (25:24:1) solution. The sample was precipitated with chilled isopropanol, washed with 70% ethanol, and air-dried. After dissolving the pellet in nuclease-free water, the DNA concentration was measured at 260 nm using a Nano-Quant Infinite 200 spectrophotometer (TECAN, Salzburg, Austria). These 652 wheat accessions were genotyped using 3.9K DArTSNP genotyping (Figure 1). Calling and filtering of DArTSNP markers

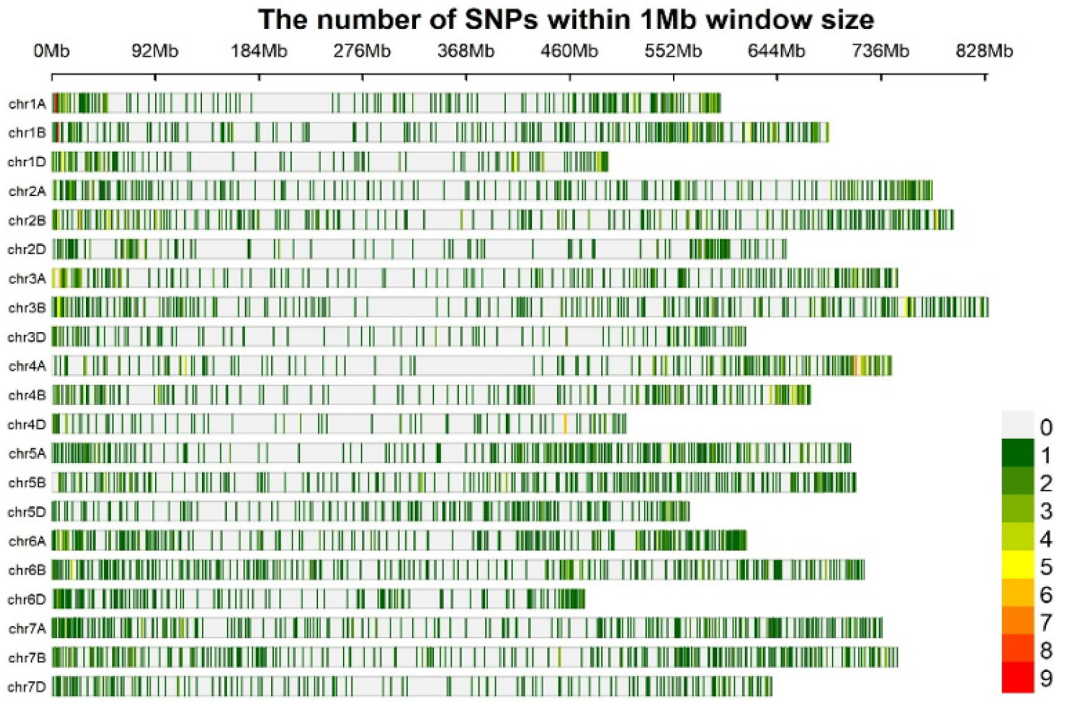


FIGURE 1
Based on their physical position, the genome-wide DArTSNP markers' density for available markers within 3.9K DArTSNP array.

were performed in the R software using the dartR package by applying call rate, reproducibility, and heterozygosity with thresholds of 0.95, 0.5, and 0.7, respectively. The DArTSNP markers' physical positions were referred to the sequences RefSeqv1.0 (https://excellenceinbreeding.org/toolbox/services/wheat-39k-mid-density-genotyping services-0).

Genome-wide association study, linkage disequilibrium, and population structure

The STRUCTURE V 2.3.4 program was used for analyzing the population structure of 652 wheat genotypes (Pritchard et al., 2000). To determine the value of K, an admixture model was employed. The value of K indicates the number of subpopulations. For every K value between 1 and 10, 10 iterations were carried out for this analysis. A 10,000 burn-in length and three independent runs for each K value were additional parameters. The STRUCTURE HARVESTER software program (Earl and VonHoldt, 2012) was used to make this analysis easier. It generates K plots for the genotypes being studied and ensures that the most optimal K value is selected. The optimal number of subpopulations (K value) was measured using the procedure described by Evanno et al. (2005).

According to Bradbury et al. (2007), the TASSEL 5 software is a reliable tool for assessing LD. To determine the associations between genetic markers and phenotypes of interest, association mapping was performed in the Rmvp package of R software using the FarmCPU model (Liu et al., 2016), general linear models (GLMs), and mixed linear models (MLMs) (Yu et al., 2006). The single-locus MLM is traditionally the most used model for GWASs. It uses population structure (Q matrix) and kinship or family relatedness (K matrix) to control spurious associations (Zhang et al., 2005; VanRaden, 2008). However, this model was designed to test one marker at a time and is more likely to cause spurious associations (Wen et al., 2018). Multi-locus model FarmCPU is considered more efficient and reliable than single-locus models for mapping studies (Vikas et al., 2022). FarmCPU operates iteratively, using both fixed and random models, and incorporates significant SNPs as cofactors in each iteration to manage spurious associations without overfitting the model (Liu et al., 2016). As stripe rust resistance is based on a complex genetic structure and wheat has a high degree of LD, the significance threshold for finding MTAs was set at −log10(p-value) > 3. These association mapping results were visualized using quantile–quantile (Q–Q) plots and Manhattan plots. Significant markers observed were further subjected to *in silico* annotation (https://plants.ensembl.org/biomart/martview). Linking the significant markers with putative candidate genes was processed using a position-dependent strategy.

Results

The genotypes for stripe rust resistance varied significantly across the locations, as shown by the ANOVA in Table 2 (p-value ≤ 0.01). The responses of several genotypes to stripe rust are depicted in Figure 2, which makes it evident how many comparable reaction types there were. According to Roelfs et al. (1992), the main infection forms have been well documented, and the modified Cobb's scale (Peterson et al., 1948) was used to calculate the severity of stripe rust as a percentage of the leaf surface area impacted. To ensure accuracy, field assessments were carried out three times. Final scores range from 80S (extremely susceptible) to 0 (immune). With a CI score between 0 and 1, a total of 148 of these genotypes exhibited a high resistance reaction at the Hisar location. Approximately 9% of the genotypes were MR, falling within the CI range of 5–15, and 327 genotypes (almost 50% of the genotypes) showed susceptible response with a CI range of 25–100. At the Karnal location, almost 44% of genotypes exhibited a moderately resistant-type reaction, with a CI value range of 5–15. Additionally, 167 genotypes (approximately 26%) showed an immune-type reaction. In the case of the Gurdaspur location, 298 genotypes (almost 46%) displayed susceptible responses with a CI value range of 25–100, and only 93 genotypes showed moderately susceptible reactions with a CI value range of 15–25. At the Khudwani location, 403 genotypes (62%) displayed moderately resistant-type reaction with a CI value range of 5–15, and 171 genotypes with a CI value range of 25–100 displayed susceptible reaction. Genotypes HI 8847, HI 8846, KBSN 17, KBSN 51, PBW 698, PBW 701, PBW 702, PBW 763, PBW 765, PBW 752, DBW 187, and WH 1270 showed highly resistant reactions across the locations. However, GABO, JW 3020, HD 2851, W 8627, BAXI 288-18, HYB-11, MACS 6145, RW 346, WR 544, and MONDHYA-32

TABLE 2 Analysis of variance of 652 wheat genotypes for stripe rust resistance.

Source	*df*	Mean sum of squares							
		Hisar		Karnal		Gurdaspur		Khudwani	
Treatment (ignoring blocks)	654	608.94	**	80.79	**	899.39	**	457.12	**
Treatment: Check	2	25.00	ns	10.71	**	24.43	ns	5.57	ns
Treatment: Test	651	601.43	**	80.46	**	884.78	**	452.97	**
Treatment: Test vs. Check	1	6,663.20	**	434.79	**	12,156.64	**	4,061.42	**
Block (eliminating Treatments)	6	6.94	ns	1.43	ns	13.11	ns	12.10	ns
Residuals	12	11.89		1.05		17.04		13.24	

**Significant at (P< 0.01).

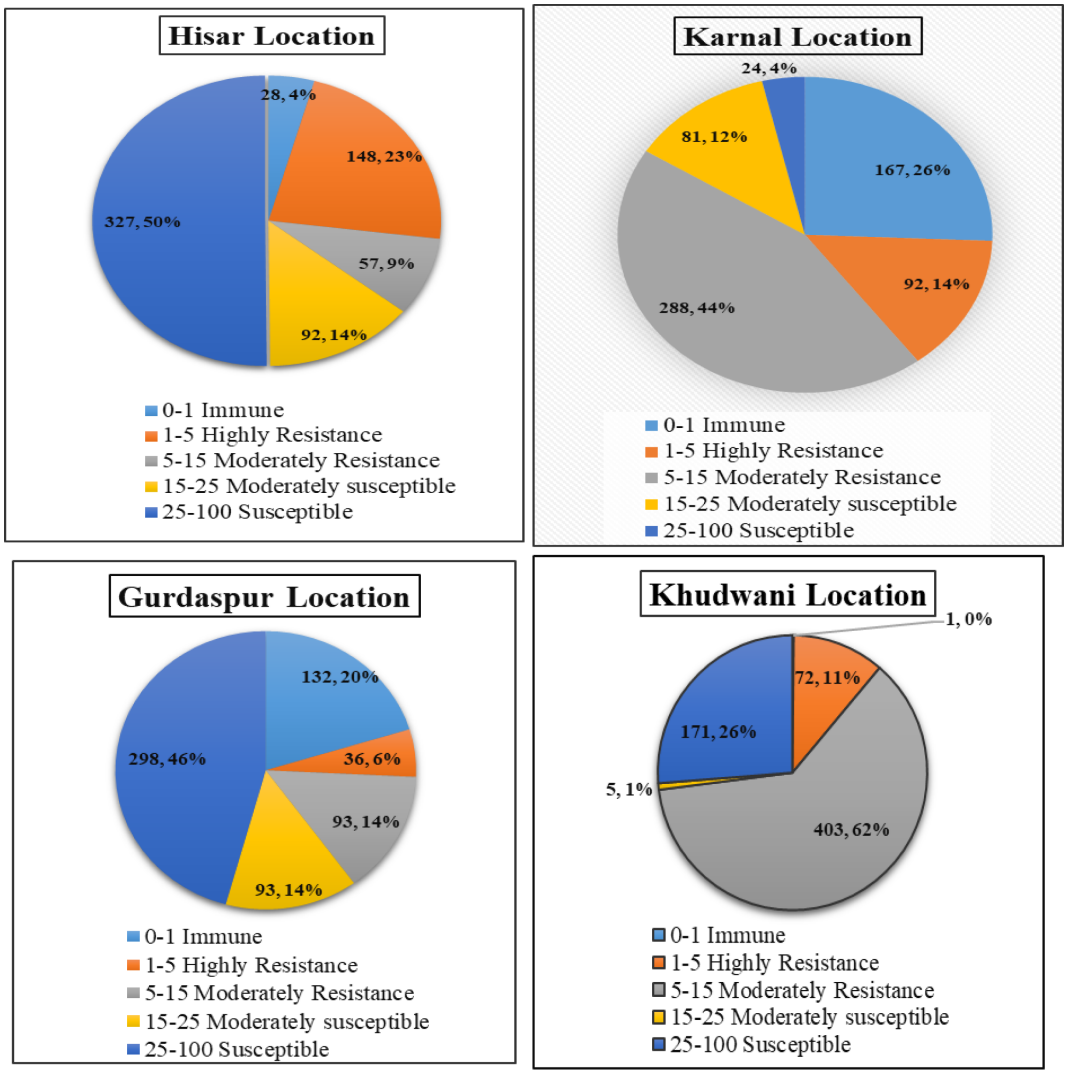


FIGURE 2
Pie chart representation of adult plant disease response against stripe rust (YR) of diverse wheat genotypes in corresponding environments (Hisar, Karnal, Gurdaspur, and Khudwani). The color legend on the right side of each pie chart represents the coefficient of infection (CI) score. The magnitude of arc length is directly proportional to the frequency of genotypes showing corresponding CI scores.

showed highly susceptible reactions across the locations. There are some genotypes, viz., Lr-13sc and NP 721, that showed resistance at Karnal and Hisar but showed susceptible reaction at the Khudwani and Gurdaspur locations. Genotype NP 825 showed resistance at Khudwani but showed a highly susceptible reaction at Hisar, Karnal, and Gurdaspur. WH 291 showed a resistant reaction at Karnal and a highly susceptible reaction at Hisar, Gurdaspur, and Khudwani. The stripe rust data were visualized and considered for studying Pearson's correlation in R statistical programming. Correlation plots for stripe rust describing the correlation of disease scores between different locations were created using the metan R package (Supplementary Figure 1). The highest correlation among disease scores was found between the Hisar and Gurdaspur locations, followed by Hisar and Khudwani. Also, some basic statistics for this data set were generated, which were indicative of variability for the trait and also the differential expression of disease across locations (Supplementary Table 5).

Analysis of molecular variance and population structure analysis

Using GenAlEx (version 6.5), Analysis of Molecular Variance (AMOVA) was performed, and it was found that the variations between groups identified by STRUCTURE analysis explained approximately 2% of the variation in the entire germplasm. Conversely, diversity within a group accounted for 98% of the variation (Figure 3). The population-wide fixation index (FST) value was 0.022, and it was deemed significant at $p < 0.001$. Variation among genotypes in the total population was particularly huge, despite pairwise FST values indicating that variation among subpopulations was lower. Data obtained from 1,938 DArTSNP markers were analyzed using the STRUCTURE software to study the population structure and for the identification of subgroups within the studied wheat genotypes.

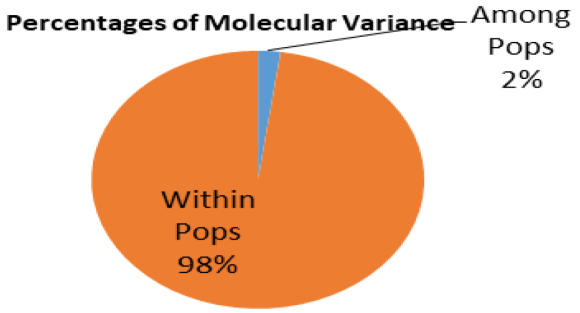


FIGURE 3
Analysis of molecular variance of subpopulation obtained through population structure analysis.

The admixture model was utilized, with a membership probability threshold of 50% established to assign genotypes to specific clusters. The results of the STRUCTURE analysis showed that there were five major subgroups within the populations, based on the ΔK values for K ranging from 1 to 10 (Figure 4). Different colors were used to demonstrate these subgroups: 238 genotypes were found in the red cluster, 182 in the green cluster, 92 in the blue cluster, 60 in the yellow cluster, and seven in the purple cluster (Figure 5). The various numbers of genotypes in each category indicated that the wheat genotypes under study had varying degrees of genetic differentiation. Additionally, 72 genotypes were categorized as admixtures, representing their genetic contributions from various subgroups, because they did not exceed the threshold membership probability for inclusion in any of the clusters.

SP1 was mainly composed of varieties (~54%), advanced breeding lines (18.6%), and mutants (~10%), while varietal lines (~70.7%) and landraces (~16.5%) collectively made up SP2. The exact contributions of SP3, SP4, and SP5 are available in Supplementary Table 2. Additionally, genetic stocks contributed ~10% in SP1, 9.3% in SP2, 18.4% in SP3, and 8.4% in SP4; landraces were observed in SP3 (29.3%); and varietal lines, which were significant components of SP1 and SP2, also appeared in SP3 with contributions of 51%. SP4 and SP5 mainly composed 55.93% and 85.7%, respectively, of varieties. The mean of Infection Type (IT) scores for genotypes in subpopulations (SP1, SP3, SP4, and SP5) was higher than that of those in subpopulation two (SP2). This shows that in the current study, genotypes were more resistant in SP2 than in SP1, SP3, SP4, and SP5 to most of the pathotypes. Therefore, the influence of population structure on rust infection was considered a covariate for subsequent association analyses.

Linkage disequilibrium

LD was studied in 94,425 locus pairings using the 1,938 DArTSNP markers from the wheat genotype panel. Among the identified locus pairings, 15,052 pairs (36.56%) exhibited significant LD, which was observed at a threshold value of $r^2 > 0.05$, indicating a non-random relationship between alleles at these loci (Figure 6). Significant LD was observed in 9,135 marker pairs when a stringent threshold of $r^2 \geq 0.1$ was applied. LD decayed at genetic distances of 7,631,202 bp for the whole genome (Figure 7). The map distance at which the fitted decay curve intersected with the critical r^2 provided an estimate of QTL-CI. The estimated QTL-CI of 76,312,02 bp was observed in this study. LD beyond this critical value is considered to be caused by genetic linkage.

Association mapping

A number of approaches, including GLM, MLM, and FarmCPU, were employed in order to find significant MTAs. The R software's Rmvp package was used for all of these approaches. Different significant MTAs were identified based on location. The picture displays quantile–quantile (QQ) graphs between the observed and expected p-values of association using GLM, MLM, and FarmCPU models, revealing how effectively these models corresponded to all four locations.

A total of 27 significant MTAs representing 14 chromosomes (Chr IA, 1B, 1D, 2B, 2D, 3B, 4A, 4D, 5A, 5D, 6A, 6B, 7A, and 7B) that were common in at least two models were found to be associated with stripe rust resistance at all four locations. At the Hisar location, out of the total identified MTAs, 14 MTAs were found that were common in at least two models to be associated with stripe rust resistance in wheat at −log10 p > 3 (Figure 8), same like in Karnal, Gurdaspur, and Khudwani, i.e., 7, 4, and 5, respectively (Figure 9). One MTA, TaDArTAG007561, was identified as common in three locations

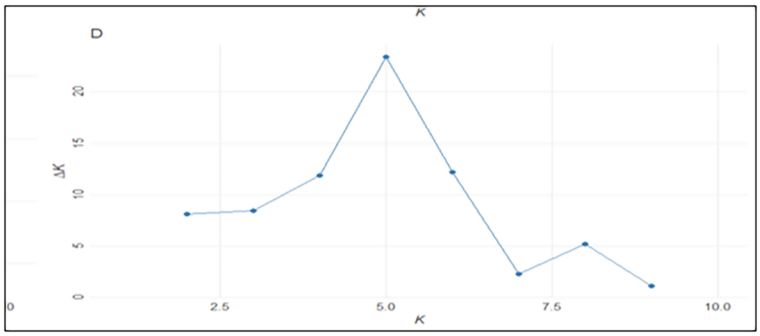


FIGURE 4
Population structure analysis: a ΔK for different numbers of subpopulations (K). Sharp peak was observed at K = 5 with maximum of ΔK.

(Figures 10, 11), i.e., Hisar, Karnal, and Khudwani; another MTA, TaDArTAG007677, was found common in two locations, i.e., Hisar and Karnal. This multi-model revealed that significant MTAs were environmentally specific, showing the presence of significant G × E interaction. The details of co-located QTLs with MTAs are given in Table 3. The emphasis was on the markers with important biological functions that have previously been validated and linked with response to disease resistance. The markers on 2B and 6B were found to correspond to leucine-rich repeat and serine/threonine-protein kinase-like domain disease resistance protein, while the SNPs identified on 6A corresponded to F-box and NB-ARC domain. Some regions also encoded an ABC transporter-like, ATP-binding domain located on 4A. Similarly, three different SNPs at chromosomes 1B, 3B, and 7A encoded receptor-like kinase proteins, which are also an important family of proteins with multiple functions, one of which is disease resistance. The many details of candidate genes are provided in Supplementary Tables 3 and Supplementary Table 4.

Discussion

The selection of highly diverse germplasm or genotypic panel is a key prerequisite for the success of any association mapping study. This diversity ensures comprehensive analysis and enhances the reliability of the findings. The statistical analysis (ANOVA) performed in the present study indicated significant variation among the genotypes. The variability for the trait and its expression across the locations can be further assessed by the basic statistics analyzed across the locations (Supplementary Table 5). Among all four regions used for screening rust resistance in the current wheat breeding program, e.g., Khudwani is considered a "yellow rust disease hotspot" in the Jammu and Kashmir regions. Also, Hisar and Karnal (Haryana) and Gurdaspur (Punjab) lie in the wheat rust-vulnerable belt. Khudwani (Jammu and Kashmir, Anantnag district) is at a higher altitude (~1,590–1,700 m above mean sea level) and lies in a Northern Hills Zone (NHZ). Gurdaspur (Punjab) lies in the sub-mountainous/pre-hill region (Kandi belt) and experiences earlier and more aggressive rust outbreaks because of its proximity to hilly inoculum sources. Hisar and Karnal (Haryana) are in the North-Western Plain Zone (NWPZ), with lower-altitude flat plains. Epidemics occur but perhaps less routinely earlier compared to those in the foothills. The onset, severity, and timing of stripe rust may differ: during December, Khudwani and Gurdaspur may face earlier and higher disease pressure because of cooler/hill-adjacent conditions; Hisar/Karnal, although susceptible, may have somewhat lower or delayed incidence depending on the season. In the current study, higher-than-average disease epidemics occur in the Gurdaspur (32.44) region, followed by Hisar (29.81166), Khudwani (19.00077), and Karnal (8.48). The possible region for lower infection at Hisar and Karnal is that the plains heat up earlier, possibly reducing the long favorable window for

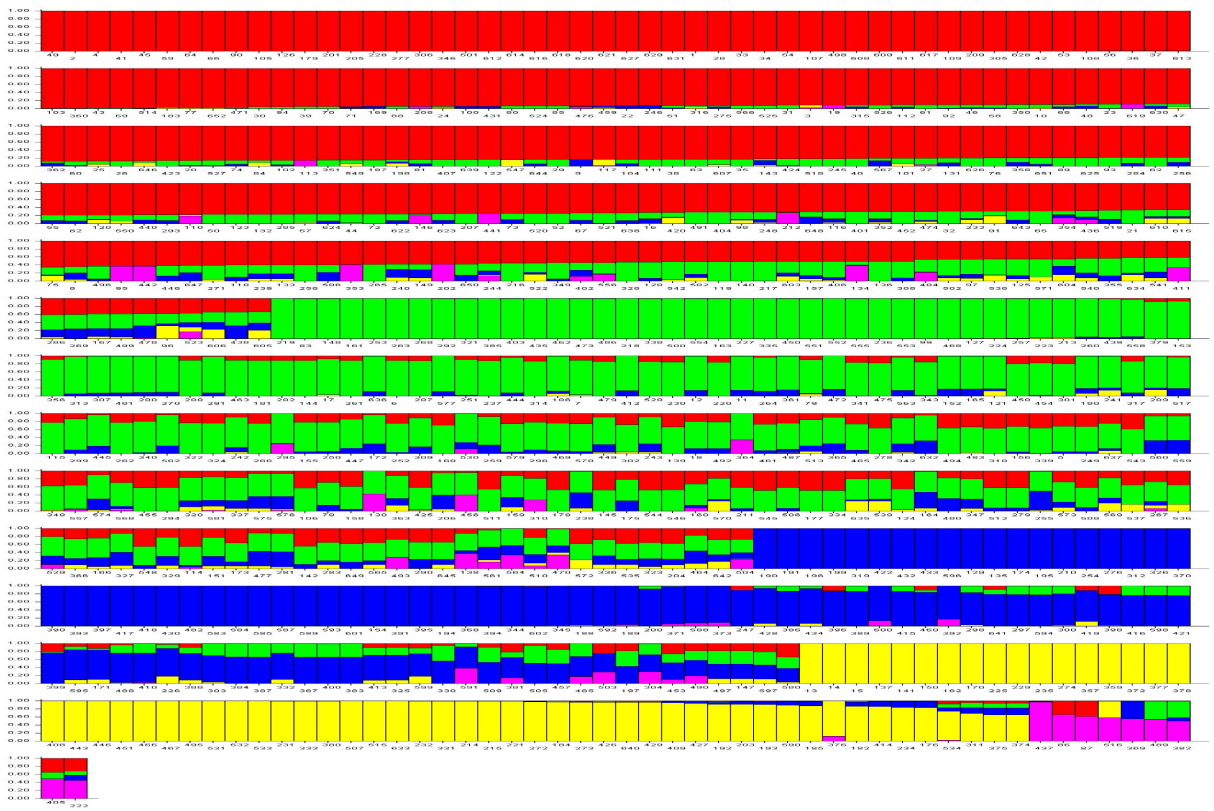


FIGURE 5
Population structure analysis: structure plot for 652 genotypes at K = 5, where each color represents one subpopulation, namely, SP1, SP2, SP3, SP4, and SP5.

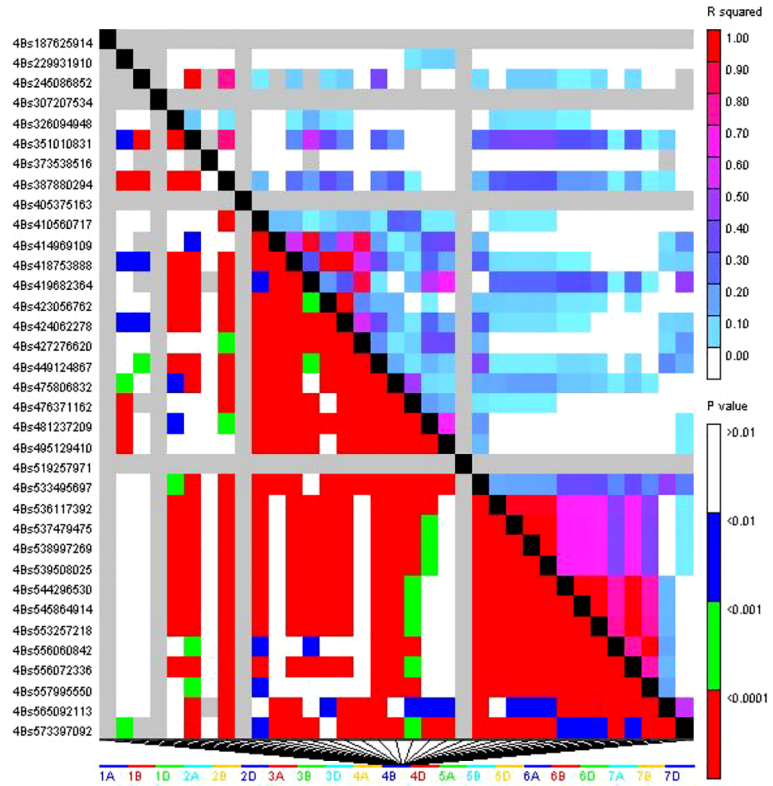


FIGURE 6
Triangle heat plot showing pairwise values of r^2 and p for different locus pairs over different chromosomes of selected wheat genotype.

rust, and maybe agronomic/varietal conditions differ, giving Hisar and Karnal somewhat lower infection. The possible reason for lower infection at Khudwani is due to slightly cooler temperatures or shorter favorable windows for rust infection.

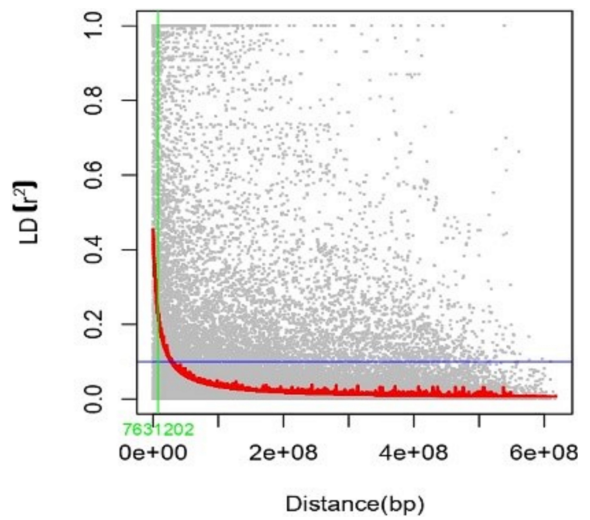


FIGURE 7
Scatter plot showing LD decay of pairwise SNP LD r2 value over the genetic distance between intra-chromosomal marker pairs.

The above results assured that this study can confidently advance with the current panel of genotypes for association mapping.

AMOVA, population structure, and LD

The AMOVA results showed that significant genetic diversity exists within subpopulations, with only 2% variation among them, and significant differences, as indicated by the partitioning value ($p < 0.001$), which is similar to an earlier study by Kumar et al. (2020). Frequent selection for economically significant features was shown by the significant variance within subpopulations. The *PhiPT* value (0.022) between subpopulations SP1, SP2, SP3, SP4, and SP5 was low, indicating little genetic differentiation among these subpopulations. This result is consistent with the AMOVA findings, which showed among-subpopulation variations accounting for only 2% of the total variation. Hence, genetic diversity analyses, including AMOVA, demonstrate that the genotypes used in this study possess ample diversity and could contribute to the development program for stripe rust resistance.

Based on the highest likelihood value of ΔK, which was identified at K = 5, the genotypes were separated into five subgroups, which were followed by K = 6 and K = 4. The relatively small number of significant clusters may be due to the close ancestral relationships and similar geographical origins of the majority of genotypes in this study. The highest likelihood score was

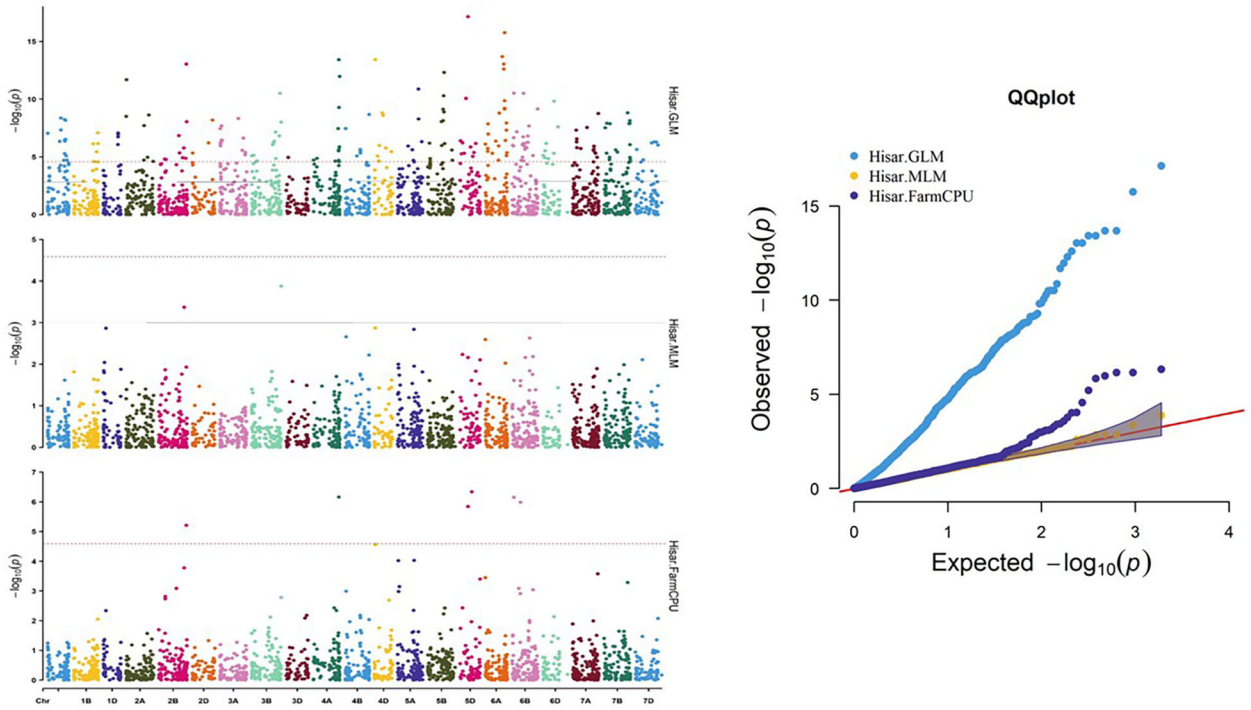


FIGURE 8
Manhattan and QQ plot for significance [–log10(p-values)] of the association of 1,938 DArTSNPs based on GLM, MLM, and FarmCPU located on 21 chromosomes with the adult plant disease responses at Hisar location. GLM, general linear model; MLM, mixed linear model.

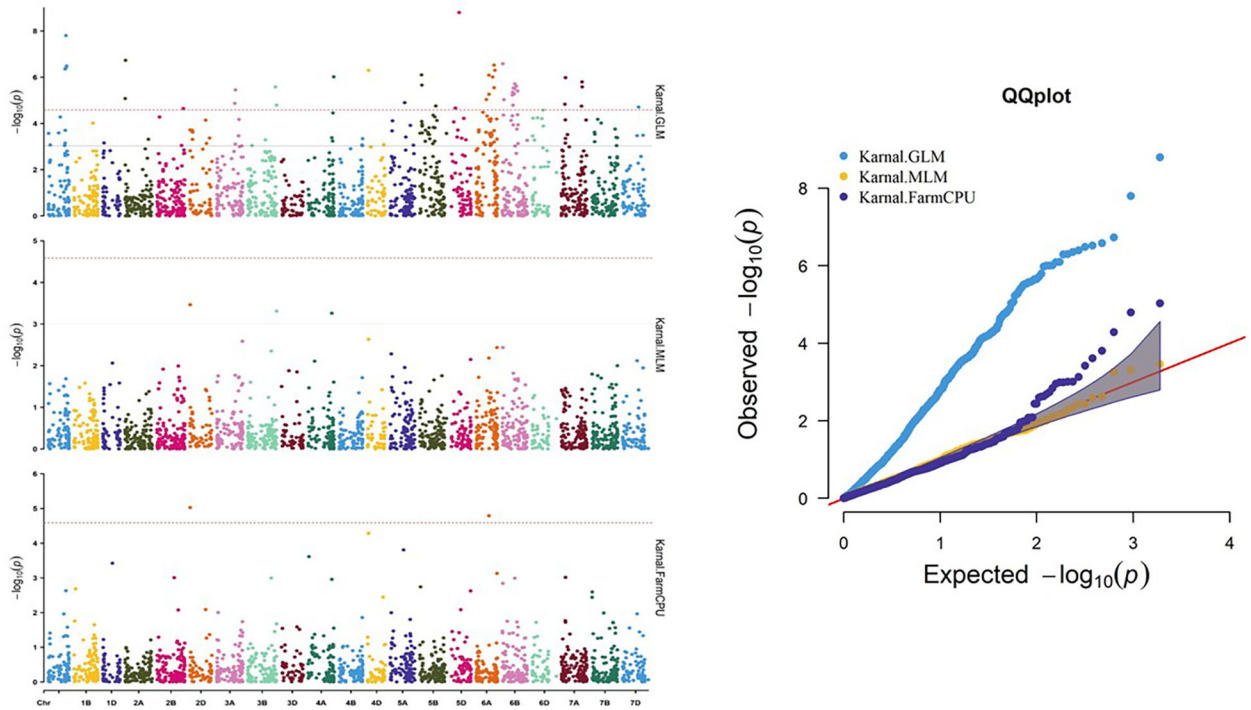


FIGURE 9
Manhattan and QQ plot for significance [–log10 (p-values)] of the association of 1,938 DArTSNPs based on GLM, MLM, and FarmCPU located on 21 chromosomes with the adult plant disease responses at Karnal location. GLM, general linear model; MLM, mixed linear model.

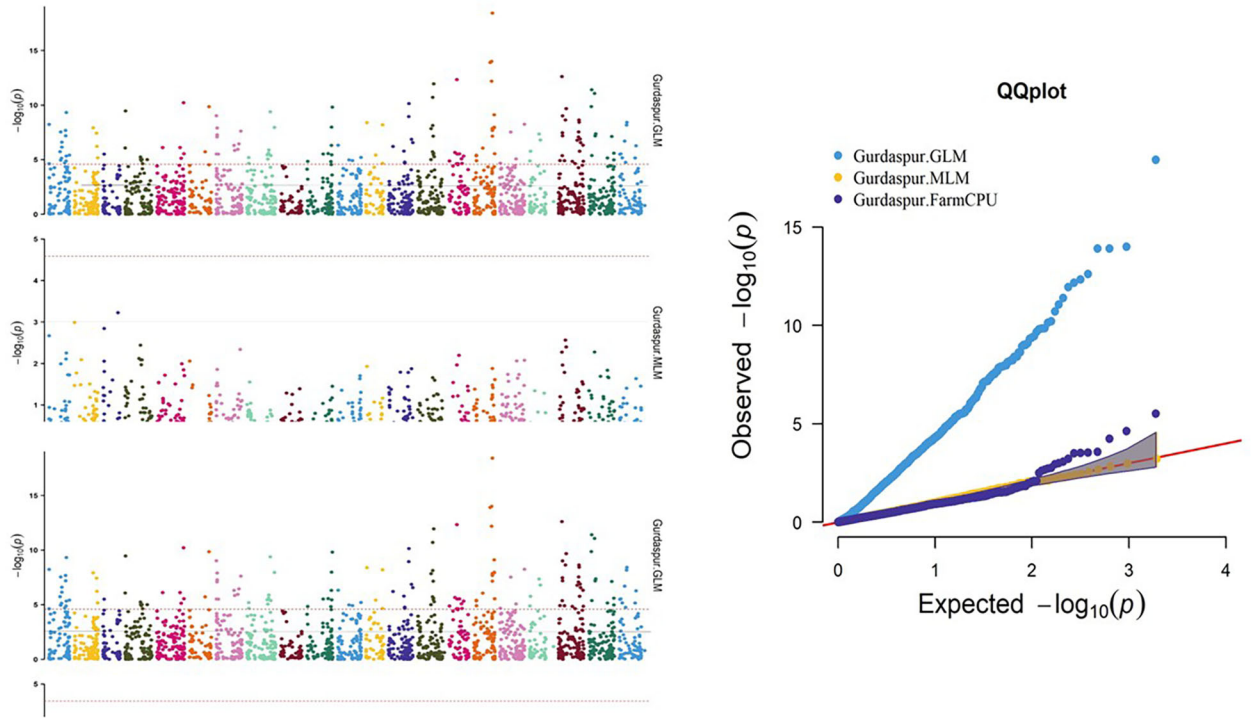


FIGURE 10
Manhattan and QQ plot for significance [–log10(p-values)] of the association of 1,938 DArTSNPs based on GLM, MLM, and FarmCPU located on 21 chromosomes with the adult plant disease responses at Gurdaspur location. GLM, general linear model; MLM, mixed linear model.

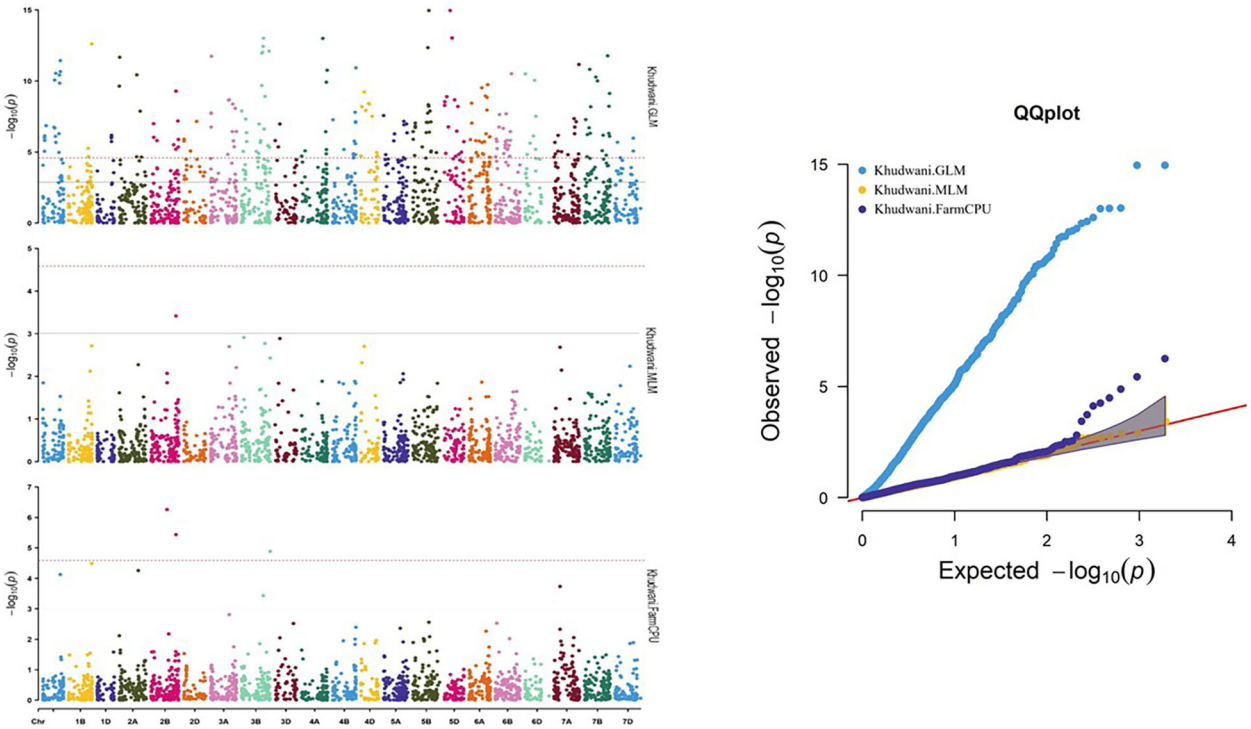


FIGURE 11
Manhattan and QQ plot for significance [–log10(p-values)] of the association of 1,938 DArTSNPs based on GLM, MLM, and FarmCPU located on 21 chromosomes with the adult plant disease responses at Khudwani location. GLM, general linear model; MLM, mixed linear model.

TABLE 3 List of MTAs identified at different locations using different models. .

Location	DArT markers	SNP name	Chromosomal location	Position	p-Value	r^2/PVE	GLM	MLM	FarmCPU	Colocalized QTLs	Reference
Karnal	TaDArTSNPSAG006173	D_contig17313_245	2D	9344681	0.000193	0.003556	P	P	P	–	–
Karnal	TaDArTSNPSAG007561	chr3B:830369163-830369463: chr3B_830369313	3B	8.3E+08	0.000153	0.003613	P	P	A	–	–
Karnal	TaDArTSNPSAG007622	chr4A:684967401-684967701: chr4A_684967551	4A	6.85E+08	0.000416	0.003066	P	P	A	*QYr.sgi-4A.1*	Agenbag et al., 2014
Karnal	TaDArTSNPSAG007769	chr6A:373995587-373995887: chr6A_373995737	6A	3.74E+08	8.22E−07	0.005706	P	A	P	*YrP10090*	Liu et al., 2021
Karnal	TaDArTSNPSAG007677	chr4D:26481348-26481648: chr4D_26481498	4D	26481498	0.000786	0.002001	P	A	P	–	–
Karnal	TaDArTSNPSAG006874	tplb0026o20_691	6A	6.07E+08	0.00048	0.002782	p	A	P	*QYr.ufs-6A*	Prins et al., 2011
Karnal	TaDArTSNPSAG007331	S7A_567030080	7A	5.67E+08	6.48E−07	0.005952	P	A	P	–	–
Hisar	TaDArTSNPSAG001883	S2B_716120567	2B	7.16E+08	0.000892	0.001949	A	P	P	*QYraq.cau-2B*, *QYrns.orz-2BL*	Guo et al., 2008; Vazquez et al., 2015
Hisar	TaDArTSNPSAG007561	chr3B:830369163-830369463: chr3B_830369313	3B	8.3E+08	0.000153	0.003613	P	P	A	–	–
Hisar	TaDArTSNPSAG000166	ALPb4A_773_SNP	4A	7.18E+08	1.07E−08	0.008659	P	A	P	–	–
Hisar	TaDArTSNPSAG004411	S6B_476413707	6B	4.76E+08	1.69E−06	0.005974	P	A	P	–	–
Hisar	TaDArTSNPSAG006947	Tdurum_contig4592_327	6B	5.72E+08	2.04E−07	0.006929	P	A	P	*QYr.nwafu-6BL*, *QYr.inra-6B*	Wu et al., 2018; Dedryver et al., 2009
Hisar	TaDArTSNPSAG009232	AX-94593432	5D	2.2E+08	4.68E−13	0.013579	P	A	P	–	–
Hisar	TaDArTSNPSAG000523	RAC875_rep_c116263_97	2B	7.75E+08	6.15E−11	0.009544	P	A	P	*QYrAvS.wgp-2BS*	Christiansen et al., 2006; Liu et al., 2020
Hisar	TaDArTSNPSAG007677	chr4D:26481348-26481648: chr4D_26481498	4D	26481498	0.000786	0.002001	P	A	P	–	–
Hisar	TaDArTSNPSAG003336	S5A_466029658	5A	4.66E+08	1.86E−05	0.003336	P	A	P	–	–
Hisar	TaDArTSNPSAG009350	AX-94979699	6A	5221698	0.000653	0.002996	P	A	P	*QYr.uga-6AS*	Hao et al., 2011
Hisar	TaDArTSNPSAG007213	BS00023166_51	7B	7.01E+08	9.07E−05	0.004002	P	A	P	–	–
Hisar	TaDArTSNPSAG003386	S5A_51793323	5A	51793323	0.000211	0.002731	P	A	P	–	–
Hisar	TaDArTSNPSAG001742	S2B_494862155	2B	4.95E+08	0.000604	0.002261	P	A	P	–	–
Hisar	TaDArTSNPSAG004467	S6B_572495855	6B	5.72E+08	0.000243	0.002822	P	A	P	–	–

(Continued)

10.3389/fpls.2025.1687331

TABLE 3 Continued

Location	DArT markers	SNP name	Chromosomal location	Position	p-Value	r^2/PVE	GLM	MLM	FarmCPU	Colocalized QTLs	Reference
Gurdaspur	TaDArTSNPSAG000018	Glu-B3fg_SNP	1B	5687022	0.000447	0.001419	P	P	P	*QYr.cim-1BS*	Zhang et al., 2022
Gurdaspur	TaDArTSNPSAG000073	TaFT3-D1	1D	4.3E+08	0.000413	0.003106	P	P	P	–	–
Gurdaspur	TaDArTSNPSAG001133	S1D_30446693	1D	30446693	3.15E–06	0.004769	P	A	P	–	–
Gurdaspur	TaDArTSNPSAG004338	S6B_229054418	6B	2.29E+08	5.14E–05	0.001602	P	A	P	–	–
Khudwani	TaDArTSNPSAG001891	S2B_717627572	2B	7.18E+08	5.44E–10	0.005772	P	P	P	*QYraq.cau-2B*, *QYrns.orz-2BL*	Guo et al., 2008; Vazquez et al., 2015
Khudwani	TaDArTSNPSAG007561	chr3B:830369163-830369463: chr3B_830369313	3B	8.3E+08	0.000153	0.003613	P	A	P	–	–
Khudwani	TaDArTSNPSAG005805	BS00066305_51	1B	6.76E+08	2.00E–13	0.010949	P	A	P	*QYr.sicau-1B.3*	Ma et al., 2019
										QYrPI181410.wgp-1BL	Liu et al., 2020
										Qyr.gaas.1B.1	Cheng et al., 2022
										QYr.crc-1BL	Rosa et al., 2019
										QYr.spa-1B	Bokore et al., 2017
										QYrsn.nwafu-1BL	Huang et al., 2021
										QYr.sicau-1BL	Wang et al., 2022
										QYrdr.wgp-1BL.2	Hou et al., 2015
										QYrsk.wgp-1BL	Liu et al., 2019
										QYrsv.swust-1BL.2	Zhou et al., 2021
										QYr.cim-1BL	Calvo-Salazar et al., 2015
Khudwani	TaDArTSNPSAG000206	S1A_500074551	1A	5E+08	5.99E–05	0.003392	P	A	P	–	–
Khudwani	TaDArTSNPSAG002689	S3B_629591171	3B	6.3E+08	0.000098	0.003964	P	A	P	–	–

MTAs, marker–trait associations; DArT, Diversity Arrays Technology; GLM, general linear model; MLM, mixed linear model; QTLs, quantitative trait loci.

10.3389/fpls.2025.1687331

reported at ΔK = 2 by Kumar et al. (2020), which resulted in the division of large germplasm into two clusters, whereas Zegeye et al. (2014); Chen et al. (2020), El-Messoadi et al. (2022), and El-Messoadi et al (2023) noted likelihood scores at ΔK = 5 or higher, which caused the clustering of genotype panels accordingly.

The results showed that there was significant linkage for 9,135 marker pairs (21.46%) at $r^2 \geq 0.1$ and for 15,056 marker pairs (36.56%) at $r^2 \geq 0.05$. The critical value of r^2 was 0.16, which was comparable to the findings of other studies by Zegeye et al. (2014); El Hanafi et al. (2021), and El-Messoadi et al (2023). Triangle plots of pairwise LD may indicate larger LD blocks due to selection pressure on genotypes or varieties within breeding programs targeted at specific desirable traits.

Association mapping and marker trait associations

Three distinct models—GLM, MLM, and FarmCPU—were used to predict the relationship between stripe rust resistance and DArTSNP markers. Several studies have shown that applying two or more distinct models to find favorable MTAs in wheat is an effective method to determine which model best fits the association mapping data (Tene et al., 2022; Abou-Zeid and Mourad, 2021; Habib et al., 2020). Comparative results from these models enhance the findings and reduce the potential for false associations. The LD results for 1,938 DArTSNP markers encouraged further association mapping. Notably, El-Messoadi et al (2022) previously conducted association mapping with 5,176 polymorphic markers using an MLM. Diversity Arrays Technology (DArT) markers on 426 synthetic bread wheat genotypes revealed three DArT markers on chromosomes 1B, 2B, and 7B that were significantly associated with resistance to stripe rust. Similarly, a comparable study conducted by El Hanafi et al. (2021) in 2014/2015 identified 23 markers that were strongly associated with adult plant resistance on chromosomes 2A, 2B, 2D, and 7B.

Thirty significant MTAs for stripe rust resistance (which were common to at least two models) were considered promising MTAs. Two markers were also found common across the locations, i.e., one in three locations and another marker in two locations, so a total of 27 genomic regions were identified. Comparative mapping with previous GWASs and QTL mapping studies for stripe rust resistance revealed that nine SNPs were co-localized within genomic regions of previously identified *Yr* genes/QTL, while 18 were located in regions not previously known to harbor stripe rust resistance genes/loci and thus were considered novel. Most of the highly significant MTAs were found on chromosomes 2B and 6B (Table 3). A low value of r^2/Proportion of Variance Explained (PVE) for stripe rust in this study may be due to the polygenic nature of resistance and also due to multi-environmental interactions between the pathogen, host, and environmental factors. Stripe rust, being a complex trait, may often lead to a small fraction of phenotypic variance due to the small effect sizes of individual loci.

Numerous *Yr* genes and QTL have been mapped on chromosome 1B, including *Yr9*, *Yr10*, *Yr15*, *Yr24/Yr26/YrCh42*, *Yr64*, *Yr65*, *YrTr1*, *YrAlp*, *YrH52*, *YrH122*, *YrL693*, *YrC142*, *YrMY41*, *QYr.cau-1BS*, *QYrco.wpg-1BS.1*, and *QYrco.wpg-1BS.2* (Feng et al., 2018). TaDArTAG000018 with SNP name Glu-B3fg_SNP was found to be associated with QYr.cim-1BS (Zhang et al., 2022), and TaDArTAG005805 with SNP name BS00066305_51 was found within the genomic regions of QYr.sicau-1B.3 (Ma et al., 2019), QYrPI181410.wgp-1BL (Liu et al., 2020b), QYr.crc-1BL (Rosa et al., 2019), QYr.spa-1B (Bokore et al., 2017), QYrsn.nwafu-1BL (Huang et al., 2021), QYr.sicau-1BL (Wang et al., 2022), QYrdr.wgp-1BL.2 (Hou et al., 2015), QYrsk.wgp-1BL, QYrsv.swust-1BL.2 (Zhou et al., 2021), and QYr.cim-1BL (Calvo-Salazar et al., 2015). BLAST analysis of these markers identified candidate genes TraesCS1B03G0005800, TraesCS1B03G0007200, and TraesCS1B03G0008000, which encode for NB-ARC, leucine-rich repeat, and serine/threonine-protein kinase domains, which play a major role in disease resistance in plants (Van Ooijen et al., 2008; Annan and Huang, 2023; Gou et al., 2015; Klymiuk et al., 2018; Kumar et al., 2020). This is a large family of resistance genes that are often directly involved in the identification of pathogen effectors and form a major component of the wheat plant immune system.

On chromosome 2B, many genes, *yr31* (Yang et al., 2013), *yr7* (Marchal et al., 2018), *yr53* (Xu et al., 2013), and *yr41* (Luo et al., 2005), have been mapped for stripe rust resistance. MTA TaDArTAG000523 with SNP name RAC875_rep_c116263_97 was found within the genomic regions of the already identified QTLs *QYraq.cau-2BL* (Guo et al., 2008) and *QYrns.orz-2BL* (Vazquez et al., 2015). BLAST analysis of these markers identified candidate genes TraesCS2B03G1451500, TraesCS2B03G1444300, and TraesCS2B03G1449600, which encode for leucine-rich repeat, protein kinase domain serine/threonine-protein kinase domains (Klymiuk et al., 2018; Gou et al., 2015), which play a major role in disease resistance in plants. The kinases are cell-surface receptors and assist in detecting the pathogen and initiating the plant's defense response. These proteins play a role during the seedling stage under higher temperatures. MTAs TaDArTAG001883, TaDArTAG001891 with SNP name *S2B_716120567*, and *S2B_717627572* were found to be associated with already identified QTL *QYrAvS.wgp-2BS* (Liu et al., 2020). BLAST analysis of these markers identified candidate genes TraesCS2B03G1299100, TraesCS2B03G1284400, and TraesCS2B03G1289400, which encode for F-box domain, protein kinase domain, and Zinc finger domain. Zinc finger protein functions as a negative regulator of programmed cell death and contributes to wheat resistance to yellow rust (Guo et al., 2013). One DArT marker, TaDArTAG001742 with SNP name *S2B*_494862155, was found far away from any QTL and already identified genes, hence tagging novel genomic regions. BLAST analysis and anchoring the flanking marker allowed the identification of the candidate gene TraesCS2B03G0888600 that encodes an ABC transporter-like, ATP-binding domain, which plays a crucial role in gene resistance (Krattinger et al., 2019). The stripe rust genes holding resistance during the adult plant stage provide broad-spectrum resistance, as they activate the general defense mechanisms.

MTA on chromosome 6B TaDArTAG004411 and TaDArTAG006947 found within two already identified QTLs *QYr.nwafu-6BL* (Wu et al., 2018) and *QYr.inra-6B* (Dedryver et al., 2009) indicated their association with stripe rust resistance. However, two other MTAs on this chromosome—TaDArTAG004338 and TaDArTAG004467—were identified away from any previously identified genes or QTLs, proving their novelty. *In silico* gene

annotation helps in the identification of candidate genes present in these genomic regions. Candidate genes TraesCS6B03G0912300 and TraesCS6B03G0914600 present in this genomic region encode serine/threonine-protein kinase domain and leucine-rich repeat. Both these domains play a major role in disease resistance against stripe rust (Gou et al., 2015; Annan and Huang, 2023). On chromosome 6A, three MTAs were identified, with all MTAs found within the genomic regions of already identified QTLs on 6A *QYr.uga-6AS* (Hao et al., 2011), *YrP10090* (Liu et al., 2021), and *QYr.ufs-6A* (Prins et al., 2011). Candidate genes identified within these genomic regions encode NB-ARC, leucine-rich repeat, disease resistance protein, plants, and F-box domain; all have disease resistance properties (Annan and Huang, 2023). Two MTAs were identified on chromosome 1D—TaDArTAG000073 and TaDArTAG001133—with SNP names TaFT3-D1 and S1D_30446693, respectively. Both of these MTAs were identified far away from previously identified genes or QTLs, showing their novelty. *In silico* gene annotation led to the identification of candidate genes. The identified candidate genes TraesCS1D03G0807200 and TraesCS1D03G0805600 encode leucine-rich repeat domain superfamily (Annan and Huang, 2023) and Zinc finger, PHD-type domain, which plays a defense role against plant disease.

Two MTAs were identified on chromosome 4A (TaDArTAG000166 and TaDArTAG007622) with SNP name *ALPb4A_773_SNP* and *chr4A:684967401-684967701:chr4A_684967551.* TaDArTAG000166 marker was found within the genomic region of QTL *QYr.sgi-4A.1* (Agenbag et al., 2014). TaDArTAG007622 marker was located very close to the already identified *yr60* gene (Herrera-Foessel et al., 2015). The candidate gene identified in the genomic region encodes a protein kinase domain, leucine-rich repeat domain, and NB-ARC domain. All these domains help in developing resistance in the plant against stripe rust. Two MTAs were identified on chromosome 5A, i.e., TaDArTAG003336 and TaDArTAG003386. TaDArTAG003336 was associated with QTL *QYr.hebau-5AL* (Gebrewahid et al., 2020), showing its role in disease resistance. TaDArTAG003386 was not found close to any loci, proving its novelty. The candidate genes (TraesCS5A03G0129300, TraesCS5A03G0129600, and TraesCS5A03G0130800) identified in the protein kinase domain, serine/threonine-protein kinase, active site, and F-box-like domain superfamily play a role in disease resistance (Chen, 2020; Brueggeman et al., 2006). MTAs identified on chromosome 3B, i.e., TaDArTAG002689 and TaDArTAG007561, were not found within the genomic regions of any genes or QTLs, showing their novelty. Candidate genes TraesCS3B03G0972900, TraesCS3B03G0973500, and TraesCS3B03G0973500 identified in their genomic regions encode domains, i.e., Zinc finger C2H2-type, leucine-rich repeat domain superfamily, and protein kinase domain, which play a defense role against plant disease (Guo et al., 2013; Annan and Huang, 2023; Chen, 2020).

Conclusion

Association mapping based on LD was successfully conducted in 652 wheat genotypes using 1,938 polymorphic DArTSNP markers. Extensive phenotyping across various locations revealed that a significant variation exists in stripe rust resistance among the genotypes. The molecular variation detected using the 1,938 polymorphic markers enables the categorization of all genotypes into five distinct clusters. Utilizing the three most effective methods for association mapping, this study identified a robust set of markers associated with stripe rust resistance. All 30 promising MTAs identified in this study corroborate findings from previous research. These identified MTAs will be instrumental for marker-assisted selection, facilitating the development of superior haplotypes with strong positive effects for stripe rust resistance and paving the way for further research on the underlying causal genes' fine mapping and cloning. The key putative candidate genes may be the important candidates for further validation and gene cloning experiments.

Data availability statement

All relevant data is contained within the article: 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

VT: Methodology, Investigation, Writing – review & editing, Software, Writing – original draft. SKu: Supervision, Methodology, Conceptualization, Writing – original draft, Resources, Writing – review & editing. CL: Project administration, Supervision, Writing – review & editing. RA: Investigation, Data curation, Writing – review & editing. RN: Data curation, Writing – review & editing, Software. DK: Writing – review & editing, Data curation. PK: Writing – review & editing, Methodology. VS: Investigation, Resources, Writing – review & editing. JS: Writing – review & editing, Methodology. SKa: Writing – review & editing, Data curation. SU: Resources, Funding acquisition, Writing – review & editing. SW: Data curation, Writing – review & editing. RS: Writing – review & editing, Supervision, Formal Analysis. RT: Supervision, Writing – review & editing, Resources.

Funding

The author(s) declare that no financial support was received for the research and/or publication of this article.

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.

Correction note

This article has been corrected with minor changes. These changes do not impact the scientific content of the article.

Generative AI statement

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

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2025.1687331/full#supplementary-material

References

Abou-Zeid, M. A., and Mourad, A. M. (2021). Genomic regions associated with stripe rust resistance against the Egyptian race revealed by genome-wide association study. *BMC Plant Biol.* 21, 42. doi: 10.1186/s12870-020-02813-6

Agenbag, G. M., Pretorius, Z. A., Boyd, L. A., Bender, C. M., MacCormack, R., and Prins, R. (2014). High-resolution mapping and new marker development for adult plant stripe rust resistance QTL in the wheat cultivar Kariega. *Mol. Breed.* 34, 2005–2020. doi: 10.1007/s11032-014-0158-4

Annan, E. N., and Huang, L. (2023). Molecular mechanisms of the Co-evolution of Wheat and Rust pathogens. *Plants* 12, 1809. doi: 10.3390/plants12091809

Aoun, M., Carter, A. H., Ward, B. P., and Morris, C. F. (2021). Genome-wide association mapping of the 'super-soft'kernel texture in white winter wheat. *Theor. Appl. Genet.* 134, 2547–2559. doi: 10.1007/s00122-021-03841-y

Aravind, J., Mukesh Sankar, S., Wankhede, D. P., and Kaur, V. (2020). "Augmented RCBD: Analysis of augmented randomised complete block designs," in *R package version 0.1* (ICAR-National Bureau of Plant Genetic Resources), 2.

Athiyannan, N., Abrouk, M., Boshoff, W. H., Cauet, S., Rodde, N., Kudrna, D., et al. (2022). Long-read genome sequencing of bread wheat facilitates disease resistance gene cloning. *Nat. Genet.* 54, 227–231.

Ballesta, P., Mora, F., and Del Pozo, A. (2019). Association mapping of drought tolerance indices in wheat: QTL-rich regions on chromosome 4A. *Scientia Agricola* 77, e20180153. doi: 10.1590/1678-992X-2018-0153

Beddow, J. M., Pardey, P. G., Chai, Y., Hurley, T. M., Kriticos, D. J., Braun, H. J., et al. (2015). Research investment implications of shifts in the global geography of wheat stripe rust. *Nat. Plants* 1, 1–5. doi: 10.1038/nplants.2015.132

Bokore, F. E., Cuthbert, R. D., Knox, R. E., Randhawa, H. S., Hiebert, C. W., DePauw, R. M., et al. (2017). Quantitative trait loci for resistance to stripe rust of wheat revealed using global field nurseries and opportunities for stacking resistance genes. *Theor. Appl. Genet.* 130, 2617–2635. doi: 10.1007/s00122-017-2980-7

Brachi, B., Morris, G. P., and Borevitz, J. O. (2011). Genome-wide association studies in plants: the missing heritability is in the field. *Genome Biol.* 12, 1–8. doi: 10.1186/gb-2011-12-10-232

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

Brueggeman, R., Drader, T., and Kleinhofs, A. (2006). The barley serine/threonine kinase gene Rpg1 providing resistance to stem rust belongs to a gene family with five other members encoding kinase domains. *Theor. Appl. Genet.* 113, 1147–1158. doi: 10.1007/s00122-006-0374-3

Calvo-Salazar, V., Singh, R. P., Huerta-Espino, J., Cruz-Izquierdo, S., Lobato-Ortiz, R., Sandoval-Islas, S., et al. (2015). Genetic analysis of resistance to leaf rust and yellow rust in spring wheat cultivar Kenya Kongoni. *Plant Dis.* 99, 1153–1160. doi: 10.1094/PDIS-07-14-0718-RE

Cheng, B., Gao, X., Cao, N., et al. (2022). QTL mapping for adult plant resistance to wheat stripe rust in M96-5 × Guixie 3 wheat population. *J Appl Genetics* 63, 265–279. doi: 10.1007/s13353-022-00686-z

Chen, X. M. (2005). Epidemiology and control of stripe rust [*Puccinia striiformis* f. sp. *tritici*] on wheat. *Can. J. Plant Pathol.* 27, 314–337. doi: 10.1080/07060660509507230

Chen, X. M. (2014).Integration of cultivar resistance and fungicide application for control of wheat stripe rust. *Can. J. Plant Pathol.* 36, 311–326.

Chen, X. (2013). High-temperature adult-plant resistance, key for sustainable control of stripe rust. *Am. J. Plant Sci.* 4, 608–627. doi: 10.4236/ajps.2013.43080

Chen, X. (2020). Pathogens which threaten food security: *Puccinia striiformis*, the wheat stripe rust pathogen. *Food Secur.* 12, 239–251. doi: 10.1007/s12571-020-01016-z

Chen, S., Rouse, M. N., Zhang, W., Zhang, X., Guo, Y., Briggs, J., et al. (2020). Wheat gene Sr60 encodes a protein with two putative kinase domains that confers resistance to stem rust. *New Phytologist* 225, 948–959.

Christiansen, M. J., Feenstra, B., Skovgaard, I. M., and Andersen, S. B. (2006). Genetic analysis of resistance to yellow rust in hexaploid wheat using a mixture model for multiple crosses. *Theor. Appl. Genet.* 112, 581–591.

Dedryver, F., Paillard, S., Mallard, S., Robert, O., Trottet, M., Negre, S., et al. (2009). Characterization of genetic components involved in durable resistance to stripe rust in the bread wheat 'Renan'. *Phytopathology* 99, 968–973. doi: 10.1094/PHYTO-99-8-0968

Dyakov, Y. T. (2007). "Phenomenology of plant-parasite relations," in *Comprehensive and molecular phytopathology* (Moscow, Russian Federation: Elsevier), 117–136.

Earl, D. A., and VonHoldt, B. M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.* 4, 359–361. doi: 10.1007/s12686-011-9548-7

El Hanafi, S., Backhaus, A., Bendaou, N., Sanchez-Garcia, M., Al-Abdallat, A., and Tadesse, W. (2021). Genome-wide association study for adult plant resistance to yellow rust in spring bread wheat (*Triticum aestivum* L.). *Euphytica* 217, 87. doi: 10.1007/s10681-021-02803-1

Ellis, J. G., Lagudah, E. S., Spielmeyer, W., and Dodds, P. N. (2014). The past, present and future of breeding rust resistant wheat. *Front. Plant Sci.* 5,p.641. doi: 10.3389/fpls.2014.00641

El-Messoadi, K., El Hanafi, S., Gataa, Z. E., Kehel, Z., Bouhouch, Y., and Tadesse, W. (2022). Genome wide association study for stripe rust resistance in spring bread wheat (Triticum aestivum L.). *J. Plant Pathol.* 104, 1049–1059. doi: 10.1007/s42161-022-01132-z

El-Messoadi, K., Rochdi, A., El-Yacoubi, H., and Wuletaw, T. (2023). *Genome wide association study for stripe rust resistance in elite spring bread wheat genotypes (*Triticum aestivum *L.) in Morocco*. Kenitra, Morocco: Biodiversity and Crop Improvement Program, ICARDA, Rabat, Morocco Faculty of Sciences Kenitra, Ibn Tofail University.

El-Messoadi, K., Rochdi, A., and Tadesse, W. (2024). Genome wide association study and genomic prediction for stripe rust resistance at the seedling stage in advanced spring bread wheat genotypes of ICARDA Morocco. *Ecol. Genet. Genomics* 31, p.100235. doi: 10.1016/j.egg.2024.100235

Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* 14, 2611–2620. doi: 10.1111/j.1365-294X.2005.02553.x

Feng, J., Wang, M., See, D. R., Chao, S., Zheng, Y., and Chen, X. (2018). Characterization of novel gene Yr79 and four additional quantitative trait loci for all-stage and high-temperature adult-plant resistance to stripe rust in spring wheat PI 182103. *Phytopathology* 108, 737–747. doi: 10.1094/PHYTO-11-17-0375-R

Flor, H. H. (1971). Current status of the genefor-gene concept. *Annu. Rev. Phytopathol.* 9, 275–296. doi: 10.1146/annurev.py.09.090171.001423

Fu, D., Uauy, C., Distelfeld, A., Blechl, A., Epstein, L., Chen, X., et al. (2009). A kinase-START gene confers temperature-dependent resistance to wheat stripe rust. *science* 323, 1357–1360. doi: 10.1126/science.1166289

Gao, L., Rouse, M. N., Mihalyov, P. D., Bulli, P., Pumphrey, M. O., and Anderson, J. A. (2017). Genetic characterization of stem rust resistance in a global spring wheat germplasm collection. *Crop Sci.* 57, 2575–2589. doi: 10.2135/cropsci2017.03.0159

Gao, Z., Wang, X., Li, Y., Hou, W., and Zhang, X. (2024). Evaluation of stripe rust resistance and genome-wide association study in wheat varieties derived from the International Center for Agricultural Research in the Dry Areas. *Front. Plant Sci.* 15, 1377253. doi: 10.3389/fpls.2024.1377253

Gebrewahid, T. W., Zhang, P., Zhou, Y., Yan, X., Xia, X., He, Z., et al. (2020). QTL mapping of adult plant resistance to stripe rust and leaf rust in a Fuyu 3/Zhengzhou 5389 wheat population. *Crop J.* 8, 655–665. doi: 10.1016/j.cj.2019.09.013

Gou, J. Y., Li, K., Wu, K., Wang, X., Lin, H., Cantu, D., et al. (2015). Wheat stripe rust resistance protein WKS1 reduces the ability of the thylakoid-associated ascorbate peroxidase to detoxify reactive oxygen species. *Plant Cell.* 27, 1755–1770. doi: 10.1105/tpc.114.134296

Gudi, S., Halladakeri, P., Singh, G., Kumar, P., Singh, S., Alwutayd, K. M., et al. (2024). Deciphering the genetic landscape of seedling drought stress tolerance in wheat (*Triticum aestivum* L.) through genome-wide association studies. *Front. Plant Sci.* 15. doi: 10.3389/fpls.2024.1351075

Gudi, S., Singh, J., Gill, H., Sehgal, S., Faris, J. D., and Gill U and Gupta, R. (2025). Phenotypic data on seedling traits of hexaploid spring wheat panel evaluated under heat stress. *Data Brief* 62, (112069). doi: 10.1016/j.dib.2025.112069

Guo, J., Bai, P., Yang, Q., Liu, F., Wang, X., Huang, L., et al. (2013). Wheat zinc finger protein TaLSD1, a negative regulator of programmed cell death, is involved in wheat resistance against stripe rust fungus. *Plant Physiol. Biochem.* 71, 164–172. doi: 10.1016/j.plaphy.2013.07.009

Guo, Q., Zhang, Z. J., Xu, Y. B., Li, G. H., Feng, J., and Zhou, Y. (2008). Quantitative trait loci for high-temperature adult-plant and slow-rusting resistance to *Puccinia striiformis* f. sp. *tritici* in wheat cultivars. *Phytopathology* 98, 803–809. doi: 10.1094/PHYTO-98-7-0803

Gururani, M. A., Venkatesh, J., Upadhyaya, C. P., Nookaraju, A., Pandey, S. K., and Park, S. W. (2012). Plant disease resistance genes: current status and future directions. *Physiol. Mol. Plant. Pathol.* 78, 51–65.

Habib, M., Awan, F. S., Sadia, B., and Zia, M. A. (2020). Genome-wide association mapping for stripe rust resistance in Pakistani spring wheat genotypes. *Plants* 9, 1056. doi: 10.3390/plants9091056

Hall, D., Tegström, C., and Ingvarsson, P. K. (2010). Using association mapping to dissect the genetic basis of complex traits in plants. *Briefings Funct. Genomics* 9, 157–165. doi: 10.1093/bfgp/elp048

Hao, Y., Chen, Z., Wang, Y., Bland, D., Buck, J., Brown-Guedira, G., et al. (2011). Characterization of a major QTL for adult plant resistance to stripe rust in US soft red winter wheat. *Theor. Appl. Genet.* 123, 1401–1411. doi: 10.1007/s00122-011-1675-8

Herrera-Foessel, S. A., Singh, R. P., Lan, C. X., Huerta-Espino, J., Calvo-Salazar, V., Bansal, U. K., et al. (2015). Yr60, a gene conferring moderate resistance to stripe rust in wheat. *Plant Dis.* 99, 508–511. doi: 10.1094/PDIS-08-14-0796-RE

Hou, L., Chen, X., Wang, M., See, D. R., Chao, S., Bulli, P., et al. (2015). Mapping a large number of QTL for durable resistance to stripe rust in winter wheat Druchamp using SSR and SNP markers. *PloS One* 10, e0126794. doi: 10.1371/journal.pone.0126794

Hovmøller, M. S., Sørensen, C. K., Walter, S., and Justesen, A. F. (2011). Diversity of Puccinia striiformis on cereals and grasses. *Annu. Rev. Phytopathol.* 49, 197–217. doi: 10.1146/annurev-phyto-072910-095230

Huang, S., Liu, S., Zhang, Y., Xie, Y., Wang, X., Jiao, H., et al. (2021). Genome-wide wheat 55K SNP-based mapping of stripe rust resistance loci in wheat cultivar Shaannong 33 and their alleles frequencies in current Chinese wheat cultivars and breeding lines. *Plant Dis.* 105, 1048–1056. doi: 10.1094/PDIS-07-20-1516-RE

Hubert, B., Rosegrant, M., van Boekel, M. A. J. S., and Ortiz, R. (2010). The future of food: scenarios for 2050. *Crop Sci.* 50, S33–S50. doi: 10.2135/cropsci2009.09.0530

Jambuthenne, D. T., Riaz, A., Athiyannan, N., Alahmad, S., Ng, W. L., Ziems, L., et al. (2022). Mining the Vavilov wheat diversity panel for new sources of adult plant resistance to stripe rust. *Theor. Appl. Genet.* 135, 1355–1373. doi: 10.1007/s00122-022-04037-8

Kang, Y., Barry, K., Cao, F., and Zhou, M. (2020). Genome-wide association mapping for adult resistance to powdery mildew in common wheat. *Mol. Biol. Rep.* 47, 1241–1256. doi: 10.1007/s11033-019-05225-4

Kankwatsa, P., Singh, D., Thomson, P. C., Babiker, E. M., Bonman, J. M., Newcomb, M., et al. (2017). Characterization and genome-wide association mapping of resistance to leaf rust, stem rust and stripe rust in a geographically diverse collection of spring wheat landraces. *Mol. Breed.* 37, 1–24. doi: 10.1007/s11032-017-0707-8

Klymiuk, V., Yaniv, E., Huang, L., Raats, D., Fatiukha, A., Chen, S., et al. (2018). Cloning of the wheat Yr15 resistance gene sheds light on the plant tandem kinase-pseudokinase family. *Nat. Commun.* 9, p.3735. doi: 10.1038/s41467-018-06138-9

Krattinger, S. G., Kang, J., Bräunlich, S., Boni, R., Chauhan, H., Selter, L. L., et al. (2019). Abscisic acid is a substrate of the ABC transporter encoded by the durable wheat disease resistance gene Lr34. *New Phytol.* 223, 853–866. doi: 10.1111/nph.15815

Krattinger, S. G., Lagudah, E. S., Spielmeyer, W., Singh, R. P., Huerta-Espino, J., McFadden, H., et al. (2009). A putative ABC transporter confers durable resistance to multiple fungal pathogens in wheat. *science* 323, 1360–1363. doi: 10.1126/science.1166453

Kumar, D., Chhokar, V., Sheoran, S., Singh, R., Sharma, P., Jaiswal, S., et al. (2020). Characterization of genetic diversity and population structure in wheat using array based SNP markers. *Mol. Biol. Rep.* 47, 293–306. doi: 10.1007/s11033-019-05132-8

Lagudah, E. S. (2011). Molecular genetics of race non-specific rust resistance in wheat. *Euphytica* 179, 81–91. doi: 10.1007/s10681-010-0336-3

Liu, L., Wang, M., Feng, J., See, D. R., and Chen, X.(2019). Whole-genome mapping of stripe rust resistance quantitative trait loci and race specificity related to resistance reduction in winter wheat cultivar Eltan. *Phytopathology*109, 1226–1235.

Liu, W., Frick, M., Huel, R., Nykiforuk, C. L., Wang, X., Gaudet, D. A., et al. (2014). The stripe rust resistance gene Yr10 encodes an evolutionary-conserved and unique CC–NBS–LRR sequence in wheat. *Mol. Plant* 7, 1740–1755. doi: 10.1093/mp/ssu112

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

Liu, S., Huang, S., Zeng, Q., Wang, X., Yu, R., Wang, Q., et al. (2021). Refined mapping of stripe rust resistance gene YrP10090 within a desirable haplotype for wheat improvement on chromosome 6A. *Theor. Appl. Genet.* 134, 2005–2021. doi: 10.1007/s00122-021-03801-6

Liu, W., Naruoka, Y., Miller, K., Garland-Campbell, K. A., and Carter, A. H. (2018). Characterizing and validating stripe rust resistance loci in US pacific northwest winter wheat accessions (*Triticum aestivum* L.) by genome-wide association and linkage mapping. *Plant Genome* 11, 170087. doi: 10.3835/plantgenome2017.10.0087

Liu, Y., Qie, Y., Li, X., Wang, M., and Chen, X. (2020). Genome-wide mapping of quantitative trait loci conferring all-stage and high-temperature adult-plant resistance to stripe rust in spring wheat landrace PI 181410. *Int. J. Mol. Sci.* 21, 478. doi: 10.3390/ijms21020478

Loegering, W. Q. (1959). *Methods for recording cereal rust data in International Spring Wheat Rust Nursery (IRN).* (USA: United States Department of Agriculture).

Luo, P. G., Ren, Z. L., Zhang, H. Q., and Zhang, H. Y. (2005). Identification, chromosome location, and diagnostic markers for a new gene (YrCN19) for resistance to wheat stripe rust. *Phytopathology®* 95, 1266–1270. doi: 10.1094/PHYTO-95-1266

Ma, J., Qin, N., Cai, B., Chen, G., Ding, P., Zhang, H., et al. (2019). Identification and validation of a novel major QTL for all-stage stripe rust resistance on 1BL in the winter wheat line 20828. *Theor. Appl. Genet.* 132, 1363–1373. doi: 10.1007/s00122-019-03283-7

Mapuranga, J., Zhang, N., Zhang, L., Liu, W., Chang, J., and Yang, W. (2022). Harnessing genetic resistance to rusts in wheat and integrated rust management methods to develop more durable resistant cultivars. *Front. Plant Sci.* 13, p.951095. doi: 10.3389/fpls.2022.951095

Marchal, C., Zhang, J., Zhang, P., Fenwick, P., Steuernagel, B., Adamski, N. M., et al. (2018). BED-domain-containing immune receptors confer diverse resistance spectra to yellow rust. *Nat. Plants* 4, 662–668. doi: 10.1038/s41477-018-0236-4

McIntosh, R. A., Dubcovsky, J., Rogers, W. J., Xia, X. C., and Raupp, W. J.(2020). V. Catalogue of Gene Symbols for Wheat: 2020 Supplement. https://wheat.pw.usda.gov/GG3/WGC.

Mihalyov, P. D., Nichols, V. A., Bulli, P., Rouse, M. N., and Pumphrey, M. O. (2017). Multi-locus mixed model analysis of stem rust resistance in winter wheat. *Plant Genome* 10, 2017–2001. doi: 10.3835/plantgenome2017.01.0001

Milus, E. A., Moon, D. E., Lee, K. D., and Mason, R. E. (2015). Race-specific adult-plant resistance in winter wheat to stripe rust and characterization of pathogen virulence patterns. *Phytopathology* 105, 1114–1122. doi: 10.1094/PHYTO-11-14-0305-R

Moore, J. W., Herrera-Foessel, S., Lan, C., Schnippenkoetter, W., Ayliffe, M., Huerta-Espino, J., et al. (2015). A recently evolved hexose transporter variant confers resistance to multiple pathogens in wheat. *Nat. Genet.* 47, 1494–1498.

Mu, J., Liu, L., Liu, Y., Wang, M., See, D. R., Han, D., et al. (2020a). Genome-wide association study and gene specific markers identified 51 genes or QTL for resistance to stripe rust in US winter wheat cultivars and breeding lines. *Front. Plant Sci.* 11, 998. doi: 10.3389/fpls.2020.00998

Muleta, K. T., Chen, X., and Pumphrey, M. (2020). Genome-wide mapping of resistance to stripe rust caused by Puccinia striiformis f. sp. tritici in hexaploid winter wheat. *Crop Sci.* 60, 115–131. doi: 10.1002/csc2.20058

Mundt, C. C. (2014). Durable resistance: a key to sustainable management of pathogens and pests. Infection. *Genet. Evol.* 27, 446–455. doi: 10.1016/j.meegid.2014.01.011

Naruoka, Y., Garland-Campbell, K. A., and Carter, A. H. (2015). Genome-wide association mapping for stripe rust (*Puccinia striiformis* f. sp. tritici) in US Pacific Northwest winter wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* 128, 1083–1101. doi: 10.1007/s00122-015-2492-2

Ni, F., Zheng, Y., Liu, X., Yu, Y., Zhang, G., Epstein, L., et al. (2023). Sequencing trait-associated mutations to clone wheat rust-resistance gene YrNAM. *Nat. Commun.* 14, p.4353. doi: 10.1038/s41467-023-39993-2

Orton, T. J. (2020). "Breeding for disease and insect resistance," in *Horticultural plant breeding* (Elsevier), 345–382.

Parlevliet, J. E., and Zadoks, J. C. (1977). The integrated concept of disease resistance: a new view including horizontal and vertical resistance in plants. *Euphytica* 26, 5–21. doi: 10.1007/BF00032062

Peterson, R. F., Campbell, A. B., and Hannah, A. E. (1948). A diagrammatic scale for estimating rust intensity on leaves and stems of cereals. *Can. J. Res.* 26, 496–500. doi: 10.1139/cjr48c-033

Prins, R., Pretorius, Z. A., Bender, C. M., and Lehmensiek, A. (2011). QTL mapping of stripe, leaf and stem rust resistance genes in a Kariega× Avocet S doubled haploid wheat population. *Mol. Breed.* 27, 259–270. doi: 10.1007/s11032-010-9428-y

Pritchard, J. K., Stephens, M., and Donnelly, P. (2000a). Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959. doi: 10.1093/genetics/155.2.945

Prospects, C. (2022).Crop Prospects and Food Situation# 4. *Food and Agriculture Organization*. doi: 10.4060/cc3233en

Qiao, L., Gao, X., Jia, Z., Liu, X., Wang, H., Kong, Y., et al. (2024). Identification of adult resistant genes to stripe rust in wheat from southwestern China based on GWAS and WGCNA analysis. *Plant Cell Rep.* 43, p.67. doi: 10.1007/s00299-024-03148-4

Riedelsheimer, C., Lisec, J., Czedik-Eysenberg, A., Sulpice, R., Flis, A., Grieder, C., et al. (2012). Genome-wide association mapping of leaf metabolic profiles for dissecting complex traits in maize. *Proc. Natl. Acad. Sci.* 109, 8872–8877. doi: 10.1073/pnas.1120813109

Risk, J. M., Selter, L. L., Krattinger, S. G., Viccars, L. A., Richardson, T. M., Buesing, G., et al. (2012). Functional variability of the Lr34 durable resistance gene in transgenic wheat. *Plant Biotechnol. J.* 10, 477–487. doi: 10.1111/j.1467-7652.2012.00683.x

Roelfs, A. P., Singh, R. P., and Saari, E. E. (1992). Rust diseases of wheat: concepts and methods of disease management. *Cimmyt.*, 81.

Rosa, S. B., Zanella, C. M., Hiebert, C. W., Brûlé-Babel, A. L., Randhawa, H. S., Shorter, S., et al. (2019). Genetic characterization of leaf and stripe rust resistance in the Brazilian wheat cultivar Toropi. *Phytopathology* 109, 1760–1768. doi: 10.1094/PHYTO-05-19-0159-R

Saghai-Maroof, M. A., Soliman, K. M., Jorgensen, R. A., and Allard, R. (1984). Ribosomal DNA spacer-length polymorphisms in barley: mendelian inheritance, chromosomal location, and population dynamics. *Proc. Natl. Acad. Sci.* 81, 8014–8018. doi: 10.1073/pnas.81.24.8014

Sánchez-Martín, J., and Keller, B. (2021). NLR immune receptors and diverse types of non-NLR proteins control race-specific resistance in Triticeae. *Curr. Opin. Plant Biol.* 62, p.102053. doi: 10.1016/j.pbi.2021.102053

Savary, S., Willocquet, L., Pethybridge, S. J., Esker, P., McRoberts, N., and Nelson, A. (2019). The global burden of pathogens and pests on major food crops. *Nat. Ecol. Evol.* 3, 430–439. doi: 10.1038/s41559-018-0793-y

Sharma, D., Avni, R., Gutierrez-Gonzalez, J., Kumar, R., Sela, H., Prusty, M. R., et al. (2024). A single NLR gene confers resistance to leaf and stripe rust in wheat. *Nat. Commun.* 15, 9925. doi: 10.1038/s41467-024-54068-6

Sharma, R., Wang, M., Chen, X., Lakkakula, I. P., Amand, P. S., Bernardo, A., et al. (2025).Genome-wide association mapping for the identification of stripe rust resistance loci in US hard winter wheat. *Theor. Appl. Genet.*138, 67.

Singh, R. P., Hodson, D. P., Jin, Y., Lagudah, E. S., Ayliffe, M. A., Bhavani, S., et al. (2015). Emergence and spread of new races of wheat stem rust fungus: continued threat to food security and prospects of genetic control. *Phytopathology* 105, 872–884. doi: 10.1094/PHYTO-01-15-0030-FI

Singh, G., Raigar, O. P., Kaur, S., Bishnoi, R., Mondal, K., Abreha, K. B., et al. (2025). Advances in genomics-assisted breeding strategies for enhancing nutrient uptake and use efficiency in cereals: A pathway toward sustainable agriculture. *Plant Stress* 18, 101002. doi: 10.1016/j.stress.2025.101002

Sørensen, C. K., Hovmøller, M. S., Leconte, M., Dedryver, F., and de Vallavieille-Pope, C. (2014). New races of Puccinia striiformis found in Europe reveal race specificity of long-term effective adult plant resistance in wheat. *Phytopathology* 104, 1042–1051. doi: 10.1094/PHYTO-12-13-0337-R

Tadesse, W., Ogbonnaya, F. C., Jighly, A., Sanchez-Garcia, M., Sohail, Q., Rajaram, S., et al. (2015). Genome-wide association mapping of yield and grain quality traits in winter wheat genotypes. *PloS One* 10, e0141339. doi: 10.1371/journal.pone.0141339

Tadesse, W., Suleiman, S., Tahir, I., Sanchez-Garcia, M., Jighly, A., Hagras, A., et al. (2019). Baum, Heat-tolerant QTLs associated with grain yield and its components in spring bread wheat under heat-stressed environments of Sudan and Egypt. *Crop Sci.* 59, 199–211. doi: 10.2135/cropsci2018.06.0389

Tene, M., Adhikari, E., Cobo, N., Jordan, K. W., Matny, O., del Blanco, I. A., et al. (2022). GWAS for stripe rust resistance in wild emmer wheat (*Triticum dicoccoides*) population: obstacles and solutions. *Crops* 2, 42–61. doi: 10.3390/crops2010005

Trethowan, R., Chatrath, R., Tiwari, R., Kumar, S., Saharan, M. S., Bains, N., et al. (2018). An analysis of wheat yield and adaptation in India. *Field Crops Res.* 219, 192–213. doi: 10.1016/j.fcr.2018.01.021

Van Ooijen, G., Mayr, G., Kasiem, M. M., Albrecht, M., Cornelissen, B. J., and Takken, F. L. (2008). Structure–function analysis of the NB-ARC domain of plant disease resistance proteins. *J. Exp. Bot.* 59, 1383–1397. doi: 10.1093/jxb/ern045

VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *J. dairy Sci.* 91, 4414–4423. doi: 10.3168/jds.2007-0980

Vazquez, M. D., Zemetra, R., Peterson, C. J., Chen, X. M., Heesacker, A., and Mundt, C. C. (2015). Multi-location wheat stripe rust QTL analysis: genetic background and epistatic interactions. *Theor. Appl. Genet.* 128, 1307–1318. doi: 10.1007/s00122-015-2507-z

Vikas, V. K., Pradhan, A. K., Budhlakoti, N., Mishra, D. C., Chandra, T., Bhardwaj, S. C., et al. (2022). Multi-locus genome-wide association studies (ML-GWAS) reveal novel genomic regions associated with seedling and adult plant stage leaf rust resistance in bread wheat (Triticum aestivum L.). *Heredity* 128, 434–449. doi: 10.1038/s41437-022-00525-1

Wan, A., and Chen, X. (2014). Virulence characterization of Puccinia striiformis f. sp. tritici using a new set of Yr single-gene line differentials in the United States in 2010. *Plant Dis.* 98, 1534–1542. doi: 10.1094/PDIS-01-14-0071-RE

Wang, Y., Hu, Y., Gong, F., Jin, Y., Xia, Y., He, Y., et al. (2022). Identification and mapping of QTL for stripe rust resistance in the Chinese wheat cultivar Shumai126. *Plant Dis.* 106, 1278–1285. doi: 10.1094/PDIS-09-21-1946-RE

Wang, H., Zou, S., Li, Y., Lin, F., and Tang, D. (2020). An ankyrin-repeat and WRKY-domain-containing immune receptor confers stripe rust resistance in wheat. *Nat. Commun.* 11, p.1353. doi: 10.1038/s41467-020-15139-6

Wen, Y. J., Zhang, H., Ni, Y. L., Huang, B., Zhang, J., Feng, J. Y., et al. (2018). Methodological implementation of mixed linear models in multi-locus genome-wide association studies. *Briefings Bioinf.* 19, 700–712. doi: 10.1093/bib/bbw145

Wu, J., Liu, S., Wang, Q., Zeng, Q., Mu, J., Huang, S., et al. (2018). Rapid identification of an adult plant stripe rust resistance gene in hexaploid wheat by high-throughput SNP array genotyping of pooled extremes. *Theor. Appl. Genet.* 131, 43–58. doi: 10.1007/s00122-017-2984-3

Xu, L. S., Wang, M. N., Cheng, P., Kang, Z. S., Hulbert, S. H., and Chen, X. M. (2013). Molecular mapping of Yr53, a new gene for stripe rust resistance in durum wheat accession PI 480148 and its transfer to common wheat. *Theor. Appl. Genet.* 126, 523–533. doi: 10.1007/s00122-012-1998-0

Yang, E. N., Rosewarne, G. M., Herrera-Foessel, S. A., Huerta-Espino, J., Tang, Z. X., Sun, C. F., et al. (2013). QTL analysis of the spring wheat "Chapio" identifies st able stripe rust resistance despite inter-continental genotype× environment interactions. *Theor. Appl. Genet.* 126, 1721–1732. doi: 10.1007/s00122-013-2087-8

Yu, L. X., Morgounov, A., Wanyera, R., Keser, M., Singh, S. K., and Sorrells, M. (2012). Identification of Ug99 stem rust resistance loci in winter wheat germplasm using genome-wide association analysis. *Theor. Appl. Genet.* 125, 749–758. doi: 10.1007/s00122-012-1867-x

Yu, J., Pressoir, G., Briggs, W. H., Vroh Bi, I., Yamasaki, M., Doebley, J. F., et al. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* 38, 203–208. doi: 10.1038/ng1702

Zadoks, J. C., Chang, T. T., and Konzak, C. F. (1974). A decimal code for the growth stages of cereals. *Weed Res.* 14, 415–421. doi: 10.1111/j.1365-3180.1974.tb01084.x

Zegeye, H., Rasheed, A., Makdis, F., Badebo, A., and Ogbonnaya, F. C. (2014). Genome-wide association mapping for seedling and adult plant resistance to stripe rust in synthetic hexaploid wheat. *PloS One* 9, e105593. doi: 10.1371/journal.pone.0105593

Zeng, Q., Wu, J., Liu, S., Chen, X., Yuan, F., Su, P., et al. (2019). Genome-wide mapping for stripe rust resistance loci in common wheat cultivar Qinnong 142. *Plant Dis.* 103, 439–447. doi: 10.1094/PDIS-05-18-0846-RE

Zhang, C., Huang, L., Zhang, H., Hao, Q., Lyu, B., Wang, M., et al. (2019). An ancestral NB-LRR with duplicated 3′ UTRs confers stripe rust resistance in wheat and barley. *Nat. Commun.* 10, p.4023. doi: 10.1038/s41467-019-11872-9

Zhang, P., Lan, C., Singh, R. P., Huerta-Espino, J., Li, Z., Lagudah, E., et al. (2022). Identification and characterization of resistance loci to wheat leaf rust and stripe rust in Afghan landrace "KU3067. *Front. Plant Sci.* 13, 894528. doi: 10.3389/fpls.2022.894528

Zhang, Y. M., Mao, Y., Xie, C., Smith, H., Luo, L., and Xu, S. (2005). Mapping quantitative trait loci using naturally occurring genetic variance among commercial inbred lines of maize (*Zea mays* L.). *Genetics* 169, 2267–2275. doi: 10.1534/genetics.104.033217

Zhang, P., Yan, X., Gebrewahid, T. W., Zhou, Y., Yang, E., Xia, X., et al. (2021). Genome-wide association mapping of leaf rust and stripe rust resistance in wheat accessions using the 90K SNP array. *Theor. Appl. Genet.* 134, 1233–1251. doi: 10.1007/s00122-021-03769-3

Zhao, K., Tung, C. W., Eizenga, G. C., Wright, M. H., Ali, M. L., Price, A. H., et al. (2011). Genome-wide association mapping reveals a rich genetic architecture of complex traits in Oryza sativa. *Nat. Commun.* 2, 467. doi: 10.1038/ncomms1467

Zhou, X., Zhong, X., Roter, J., Li, X., Yao, Q., Yan, J., et al. (2021). Genome-wide mapping of loci for adult-plant resistance to stripe rust in durum wheat Svevo using the 90K SNP array. *Plant Dis.* 105, 879–888. doi: 10.1094/PDIS-09-20-1933-RE

Zhu, Z., Cao, Q., Han, D., Wu, J., Wu, L., Tong, J., et al. (2023). Molecular characterization and validation of adult-plant stripe rust resistance gene Yr86 in Chinese wheat cultivar Zhongmai 895. *Theor. Appl. Genet.* 136, 142. doi: 10.1007/s00122-023-04374-2

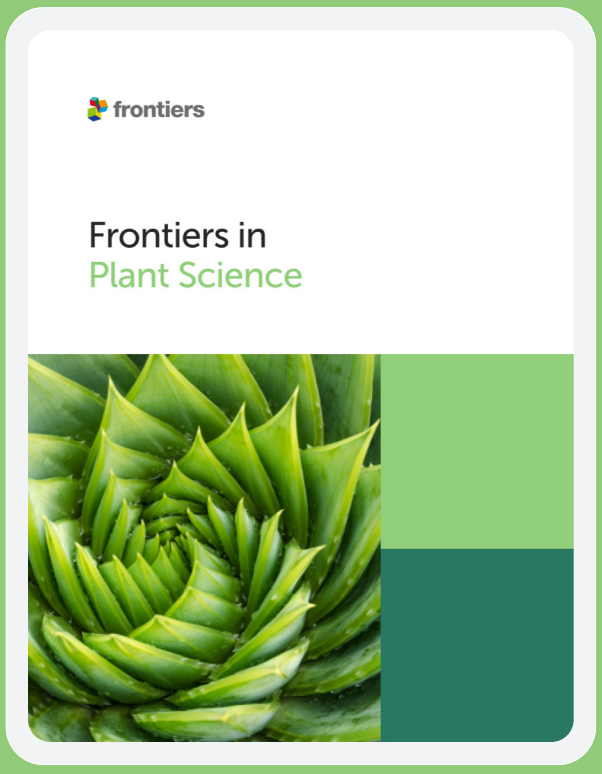
Frontiers in
Plant Science

Cultivates the science of plant biology and its applications

The most cited plant science journal, which advances our understanding of plant biology for sustainable food security, functional ecosystems and human health.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact