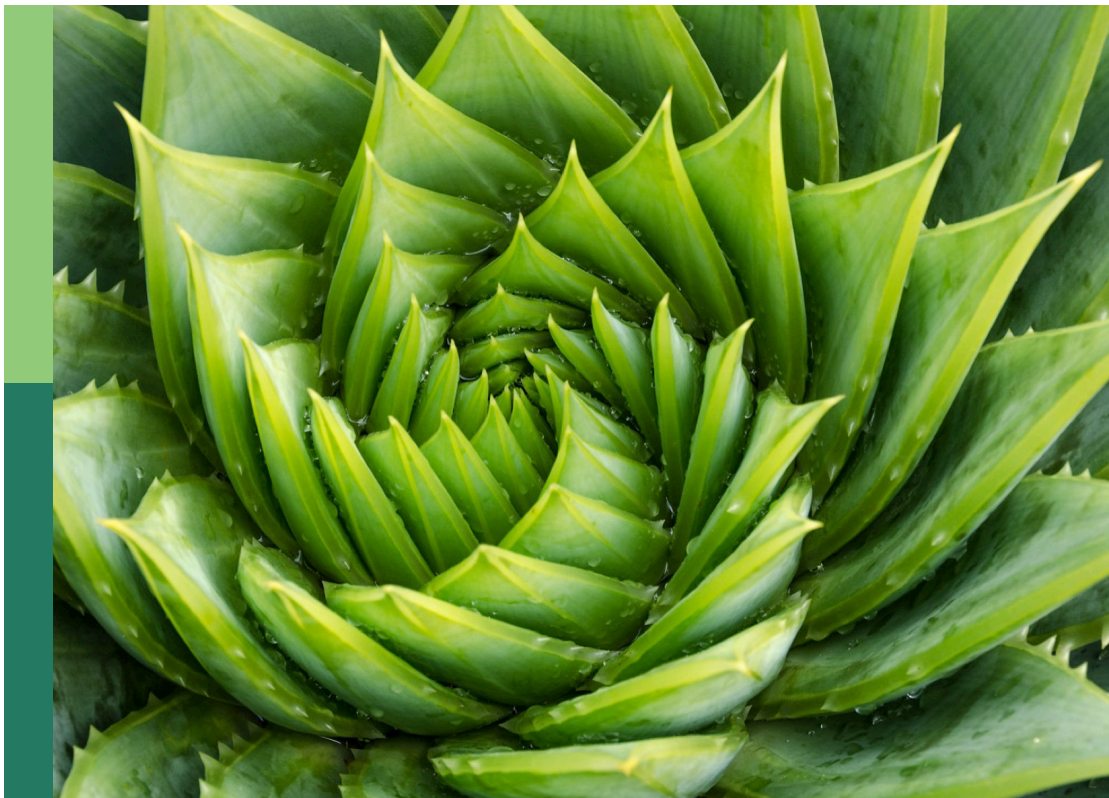
frontiers | Research Topics

Genetics and genomics of emerging and multifactorial stresses affecting plant survival and associated plant microbiomes

Edited by
Somashekhar M. Punnuri, Mahendar Thudi and Reyazul Rouf Mir

Coordinated by
Susan Moenga

Published in
Frontiers in Plant Science



FRONTIERS EBOOK COPYRIGHT STATEMENT

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

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

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

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

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

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

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

ISSN 1664-8714
ISBN 978-2-8325-7341-9
DOI 10.3389/978-2-8325-7341-9

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

Genetics and genomics of emerging and multifactorial stresses affecting plant survival and associated plant microbiomes

Topic editors

Somashekhar M. Punnuri — Fort Valley State University, United States

Mahendar Thudi — Fort Valley State University, United States

Reyazul Rouf Mir — Sher-e-Kashmir University of Agricultural Sciences and Technology of Kashmir, India

Topic coordinator

Susan Moenga — International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Kenya

Citation

Punnuri, S. M., Thudi, M., Mir, R. R., Moenga, S., eds. (2026). *Genetics and genomics of emerging and multifactorial stresses affecting plant survival and associated plant microbiomes*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-7341-9

Table of contents

06 Editorial: Genetics and genomics of emerging and multifactorial stresses affecting plant survival and associated plant microbiomes
Somashekhar M. Punnuri, Mahendar Thudi and Reyazul Rouf Mir

11 Genome-wide identification and expression analysis of protein arginine methyltransferase and JmjC domain-containing family in apple
Shenghui Su, Min Ji, Jiaqi Chen, Meidie Zhang, Xiaozhao Xu and Chenxia Cheng

29 Uncovering the effect of waterlogging stress on plant microbiome and disease development: current knowledge and future perspectives
Anshika Tyagi, Sajad Ali, Rakeeb Ahmad Mir, Sandhya Sharma, Kumari Arpita, Mohammed A. Almalki and Zahoor Ahmad Mir

39 Differential expression of microRNAs in response to *Papaya ringspot virus* infection in differentially responding genotypes of papaya (*Carica papaya* L.) and its wild relative
Basavaprabhu L. Patil and Savarni Tripathi

55 Exploring the genetic architecture of powdery mildew resistance in wheat through QTL meta-analysis
Divya Sharma, Neeraj Budhlakoti, Anita Kumari, Dinesh Kumar Saini, Anshu Sharma, Aakash Yadav, Reyazul Rouf Mir, Amit Kumar Singh, V. K. Vikas, Gyanendra Pratap Singh and Sundeep Kumar

69 Divergent evolution of *NLR* genes in the genus *Glycine*: impacts of annuals and perennials' life history strategies
Abu Bakar Sultan, Humera Nawaz, Fozia Saleem, Sehar Nawaz, Muhammad Danial, Romana Iftikhar, Umer Maqsood, Amna Areej, Sidra Shakoor, Nada H. Aljarba, Rizwan Maqbool, Muhammad Rizwan and Saad Serfraz

83 Date palm transcriptome analysis provides new insights on changes in response to high salt stress of colonized roots with the endophytic fungus *Piriformospora indica*
Manzoor Ahmad, Mughair Abdul Aziz, Miloofer Sabeem, M. Sangeeta Kutty, Sathesh K. Sivasankaran, Faical Brini, Ting Ting Xiao, Ikram Blilou and Khaled Masmoudi

97 Genetic dissection of value-added quality traits and agronomic parameters through genome-wide association mapping in bread wheat (*T. aestivum* L.)
Manish K. Vishwakarma, Pradeep K. Bhati, Uttam Kumar, Ravi P. Singh, Sundeep Kumar, Velu Govindan, Gurvinder Singh Mavi, Karthikeyan Thiyagarajan, Narain Dhar and Arun K. Joshi

117 Genome-wide association mapping identifies novel SNPs for root nodulation and agronomic traits in chickpea
B. S. Chandana, Rohit Kumar Mahto, Rajesh Kumar Singh, Aditi Bhandari, Gitanjali Tandon, K. K. Singh, Sunita Kushwah, Gera Roopa Lavanya, Mir Asif Iquebal, Neelu Jain, Himabindu Kudapa, H. D. Upadhyaya, Aladdin Hamwieh and Rajendra Kumar

136 Characterization of rhizosphere bacterial communities in oilseed rape cultivars with different susceptibility to *Plasmodiophora brassicae* infection
Yue Deng, Wenxian Wu, Xiaoqing Huang, Xiaoxiang Yang, Yaoyin Yu, Zhongmei Zhang, Zijin Hu, Xiquan Zhou, Kang Zhou, Yong Liu and Lei Zhang

150 Inheritance of resistance to maize lethal necrosis in tropical maize inbred lines
Hilda M. Kavai, Dan Makumbi, Felister M. Nzuve, Vincent W. Woyengo, L. M. Suresh, William M. Muiru, Manje Gowda and Boddupalli M. Prasanna

166 Profiling of rhizosphere-associated microbial communities in North Alabama soils infested with varied levels of reniform nematodes
Sowndarya Karapareddy, Varsha C. Anche, Sowjanya R. Tamatamu, Madhusudhana R. Janga, Kathy Lawrence, Leopold M. Nyochembeng, Antonette Todd, Lloyd T. Walker and Venkateswara R. Sripathi

184 Use of multi-trait principal component selection index to identify fall armyworm (*Spodoptera frugiperda*) resistant maize genotypes
Wilber Wambi, Dan Makumbi, Godfrey Asea, Habtamu Zeleke, Anani Y. Bruce, Mulatu Wakgari, Daniel Bomet Kwemoi and Boddupalli M. Prasanna

199 Identification of gray leaf spot–resistant donor lines in tropical maize germplasm and their agronomic performance under artificial inoculation
L. M. Suresh, Manje Gowda, Yoseph Beyene, Dan Makumbi, Kulai Amadu Manigben, Juan Burgueño, Robert Okayo, Vincent W. Woyengo and Boddupalli M. Prasanna

218 Consensus genomic regions and key genes for biotic, abiotic and key nutritional traits identified using meta- QTL analysis in peanut
Aakash Sahu, Sagar Krushnaji Rangari, Yogesh Dashrath Naik, Anjali Jyotish, Manish K. Pandey, Rajeev K. Varshney, Mahendar Thudi and Somashekhar M. Punnuri

231 Comprehensive metabolomic and microbial analysis of tobacco rhizosphere soil responses to crop rotation and fertilization
Xinyu Liu, Ying Qiao, Xianzhi Wu, Xuanxuan Chen, Fan Yang, Hao Li, Chuanzong Li, Yong Yang, Chunlei Yang, Jun Yu and Pan Luo

249 Mapping QTLs for *Pyricularia* leaf spot, nematode resistance, and yield related traits in pearl millet [*Cenchrus americanus* (L.) Morrone]
Sairam Vutla, Joseph E. Knoll, Anvesh Sankuratri, Risha G. Nayak, Limei Liu, Peng W. Chee, Raghupathy Karthikeyan, Bashasab Fakrudin, Mahendar Thudi, Patricia Timper, Karen Harris-Shultz, Jason G. Wallace, Hari Singh, Bharat Singh and Somashekhar M. Punnuri

263 Unveiling the mechanisms driving the rapid growth of *Malania oleifera* seedlings, a high-value root hemiparasitic plant
Si-Hai Wang, Chuan-Guang Zhang, Wei Yang, Jian Chen and Ming Shi

frontiers | Frontiers in Plant Science

TYPE Editorial
PUBLISHED 17 December 2025
DOI 10.3389/fpls.2025.1738816

Check for updates

OPEN ACCESS

EDITED AND REVIEWED BY
Khurram Bashir,
Lahore University of Management Sciences, Pakistan

*CORRESPONDENCE
Somashekhar M. Punnuri
punnuris@fvsu.edu

RECEIVED 03 November 2025
ACCEPTED 01 December 2025
PUBLISHED 17 December 2025

CITATION
Punnuri SM, Thudi M and Mir RR (2025) Editorial: Genetics and genomics of emerging and multifactorial stresses affecting plant survival and associated plant microbiomes. *Front. Plant Sci.* 16:1738816. doi: 10.3389/fpls.2025.1738816

COPYRIGHT
© 2025 Punnuri, Thudi and Mir. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Editorial: Genetics and genomics of emerging and multifactorial stresses affecting plant survival and associated plant microbiomes

Somashekhar M. Punnuri[1]*, Mahendar Thudi[1] and Reyazul Rouf Mir[2]

[1]Agriculture Research Station, College of Agriculture Family Science and Technology (CAFST), Fort Valley State University, Fort Valley, GA, United States, [2]Department of Genetics and Plant Breeding, Sher-e-Kashmir University of Agricultural Sciences & Technology of Kashmir (SKUAST-K), Srinagar, Kashmir, India

KEYWORDS

genome-wide association studies (GWAS), Transcriptomics, QTL mapping, epigenetic profiling, functional genomics, microbiome research, stress resilience, holobiont

Editorial on the Research Topic
Genetics and genomics of emerging and multifactorial stresses affecting plant survival and associated plant microbiomes

In the face of unprecedented global climate challenges, agricultural systems must adapt to a convergence of multifactorial stresses—drought, salinity, temperature fluctuations, nutrient limitations, pathogen outbreaks and other abiotic stresses—many of which occur concurrently, amplifying their impact and threatening crop yield, quality, and global food security (Fedoroff et al., 2010; Jiang et al., 2025). These multifactorial stressors not only threaten global crop yields and food security but also expose the underexplored but critical role of plant-associated microbiomes in mediating plant health, development, and resilience. This complex biological interplay provides the backdrop for the Research Topic *"Genetics and Genomics of Emerging and Multifactorial Stresses Affecting Plant Survival and Associated Plant Microbiomes"*.

Bringing together sixteen original research and one mini-review articles, this Research Topic presents a multifaceted exploration into how plants perceive, respond to, and adapt under layered stress conditions. Through the integration of genome-wide analyses such as Genome-Wide Association Studies (GWAS), transcriptomics, QTL mapping, epigenetic profiling, functional genomics and microbiome research, these studies not only deepen our understanding of plant stress biology but also lay the groundwork for a paradigm shift in crop improvement—one that embraces holobiont-based breeding (Huitzil et al., 2023) and systems-level thinking for resilience in a rapidly changing climate.

Advancing genetic dissection through genetic mapping, GWAS and QTL meta-analysis

Several contributions leverage powerful genomics tools to identify genes and markers associated with stress-related traits. Vutla et al. dissected the genetic basis of eight key traits in pearl millet (*Cenchrus americanus*) using a recombinant inbred line population and high-density SNP map, identifying 45 QTLs. The co-localization of multiple QTLs on LG3 and consistent detection across years emphasizes the robustness and breeding relevance for the yield-related traits and also suggesting linked improvement of traits like plant height and panicle size in pearl millet. The overlap of QTLs means breeders can potentially improve multiple traits with fewer selection cycles. This work helps secure millet's role as a hardy cereal for food and fodder in marginal lands. Sahu et al. conducted a comprehensive meta-analysis consolidating QTL data from 30 studies over 12 years, leading to the identification of 70 high-confidence meta-QTLs associated with yield, stress tolerance, and aflatoxin resistance in peanut (*Arachis hypogaea*). The discovery of candidate genes linked to aflatoxin resistance and fatty acid composition offers direct targets for marker-assisted programs. This sets the stage for peanuts that are safer, healthier, and more climate-resilient. Chandana et al. conducted a large-scale genome-wide association studies (GWAS) in chickpea (*Cicer arietinum*), identifying over 1,000 marker–trait associations, including 75 novel loci related to root nodulation. Their work provides critical genetic insights into enhancing biological nitrogen fixation, an essential trait for sustainable crop production. In wheat, Sharma et al. employed QTL meta-analysis to consolidate over 200 QTLs associated with powdery mildew resistance into 68 meta-QTLs and 13 high-confidence MQTLs, some co-localizing with known resistance genes in wheat. The refined loci enable fine mapping and functional studies to enhance durable resistance. Complementarily, Vishwakarma et al. performed GWAS to map SNPs linked to grain quality and agronomic traits in bread wheat (*Triticum aestivum*). Their identification of stable, environment-resilient SNPs informs breeding programs aimed for dual improvement in yield and quality. Suresh et al. screened 427 tropical maize lines and identified 14 lines with robust Gray Leaf Spot resistance, some carrying extra drought or viral resistance. The genetic markers found provide a toolkit for breeders to combine disease resistance with agronomic performance. Given GLS's yield impact in Africa, these donor lines could substantially reduce crop losses without heavy fungicide use. Together, these studies illustrate the power of modern genomic tools to untangle complex stress-related traits and move toward the development of crop varieties that are not only high-yielding but also resilient to multifactorial environmental stresses. Their collective contributions mark a critical step in incorporating multi-trait and environment-stable genomic regions into mainstream breeding programs.

Molecular, epigenetic, and evolutionary insights of gene families in plant defense and disease resistance

Understanding gene expression and regulatory mechanisms under stress is essential for designing resilient crops. Ahmad et al. explored the transcriptomic response of date palm roots to salinity stress in the presence of the beneficial root endophyte *Piriformospora indica*. Their study revealed upregulation of genes involved in ion transport, oxidative stress responses, and hormone signaling highlighting the synergistic roles of beneficial microbes in stress mitigation. Liu et al. conducted a comprehensive metabolomic and microbial profiling of tobacco rhizosphere soils across four cropping systems, using non-targeted metabolomics and amplicon sequencing (16S rRNA and ITS). Their findings revealed significant shifts in lipid metabolism, amino acid biosynthesis, and secondary metabolite pathways, which in turn shaped distinct microbial communities. Notably, increased abundance of arbuscular mycorrhizal and saprotrophic fungi suggests enhanced nutrient cycling and plant support under diversified cropping. This study demonstrates the intricate link between soil chemistry and microbial dynamics and provides actionable insights for designing sustainable rotation and fertilization strategies in tobacco cultivation. Patil and Tripathi profiled microRNA expressions in papaya (*Carica papaya*) genotypes with contrasting responses to Papaya ringspot virus (PRSV), revealing genotype-specific and infection-responsive miRNA signatures. Their findings underscore the regulatory role of small RNAs in modulating antiviral defenses and highlight miRNA-based mechanisms that differentiate susceptible and resistant responses. This study opens avenues for molecular breeding and genome-editing approaches to enhance PRSV resistance in papaya. Su et al. provided a comprehensive analysis of Protein Arginine Methyltransferase (PRMT) and Jumonji C-domain containing (JmjC) gene families in apple (*Pyrus malus*), linking their expression to cold and drought stress. As key players in histone modification and epigenetic regulation, these genes are promising candidates for engineering stress-responsive chromatin dynamics in perennial crops. Adding an evolutionary dimension, Sultan et al. studied NLR immune gene evolution in annual and perennial Glycine species. Annuals like soybeans had expanded NLRomes via recent duplications, while perennials experienced contraction but higher diversification. Gene birth after speciation contributed to unique repertoires, especially in *G. latifolia*. Uneven NLR distribution was observed in polyploid genomes. Findings inform disease resistance breeding in soybean. This study underscores the value of leveraging wild relatives for breeding programs and highlights the broader importance of germplasm diversity in anticipating future pathogen pressures. Evolutionary divergence and gene duplication events have shaped plant immunity-related gene families (e.g., NLRs, PRMTs, JMJs),

offering novel genetic resources for biotechnological and breeding interventions.

Collectively, these insights reveal the sophisticated regulatory frameworks plants employ to navigate environmental stress and emphasize untapped avenues for breeding resilient cultivars through the integration of molecular and evolutionary principles.

Functional genomics and field validation

Several studies bridged molecular discoveries with practical applications. Wambi et al. employed a multi-trait, principal component–based selection index on 192 maize hybrids to identify genotypes resistant to fall armyworm (FAW). The best index improved yield under infestation while cutting leaf damage significantly. This approach lets breeders target multiple traits at once, speeding development of pest-resilient hybrids. Such tools could be crucial for African farmers battling FAW without heavy pesticide reliance. This approach offers a robust and efficient tool for breeding high-yielding, FAW-resistant maize varieties. Kavai et al. investigated the genetic basis of resistance to maize lethal necrosis (MLN) in tropical maize by evaluating 182 hybrids from a 14-parent diallel across three years under artificial inoculation and rainfed conditions in Kenya. Identifying inbred lines with both resistance and yield potential allows for hybrid development without major trade-offs. In regions hit by MLN, these results offer a pathway to stable maize production in sub-Saharan Africa. Wang et al. offer a mechanistic foundation for hemiparasitic seedling development in *Malania oleifera*, an ecologically significant and oil-rich tree endemic to karst regions. Growth trials with nutrient-rich/poor soils and various hosts showed vigorous hosts greatly improved aboveground growth, with less effect on roots. Hormone metabolism, stress response, and antibiotic biosynthesis genes were upregulated in haustoria. Host association boosted nutrient synthesis and stress tolerance. Findings aid cultivation of hard-to-grow hemiparasites while optimizing propagation strategies for this economically valuable species.

Plant-microbiome interactions under multifactorial stresses

An essential component of this Research Topic is the examination of plant-microbiome interactions in stress contexts: Rhizosphere microbial communities are dynamic mediators of plant stress responses and disease resistance. Crop genotype, soil conditions, and biotic stresses co-influence microbial shifts, with strong implications for sustainable agriculture.

Liu et al. studied tobacco cropping systems, showing that crop rotation and fertilization alter rhizosphere metabolites (lipids, amino acids) and microbial diversity (e.g., mycorrhizae), enhancing soil health and plant productivity. Tyagi et al. provide a timely mini-review on the complex interplay between waterlogging stress, plant microbiomes, and disease development. The authors emphasized how waterlogging induces metabolic reprogramming, hypoxia, nutrient imbalances, and shifts in microbial community structure, all of which can exacerbate pathogen incidence and compromise plant resilience. This synthesis identifies critical knowledge gaps and lays the groundwork for integrating microbial ecology into waterlogging-tolerant crop management strategies under climate change scenarios. Karapareddy et al. profiled rhizosphere microbial communities in cotton-growing soils across North Alabama with varied levels of reniform nematode infestation using 16S and ITS amplicon sequencing. Their study identified over 47,000 bacterial and 3,400 fungal ASVs, with key bacterial genera such as *Bacillus*, *Streptomyces*, and *Conexibacter*, and fungal genera including *Fusarium* and *Cladosporium*. The community structure showed tight clustering among Actinobacteria, Acidobacteria, and Proteobacteria, suggesting functional synergy. These findings underscore the ecological relevance of microbial diversity in modulating plant–nematode interactions and provide a foundation for rhizosphere-targeted strategies in pest and soil health management. Deng et al. characterized rhizosphere bacterial communities in oilseed rape (*Brassica napus*) cultivars with contrasting responses to *Plasmodiophora brassicae* infection, the causal agent of clubroot disease. Using amplicon sequencing and metagenomic functional analysis, the study revealed that resistant and susceptible cultivars exhibited distinct shifts in key nitrogen-cycling bacterial genera—such as *Nitrosomonas*, *Limnobacter*, and *Thiobacillus*—under pathogen stress. Notably, susceptible cultivars displayed enhanced bacterial co-occurrence network complexity and upregulation of nitrification genes, while resistant cultivars favored assimilatory nitrate reduction pathways. These findings emphasize the dynamic role of rhizosphere microbiomes in modulating host-pathogen interactions and provide insight into microbial contributions to disease resistance mechanisms. These studies collectively advance the plant–microbiome integration paradigm—the concept that plant resilience emerges from the co-functioning of plant and microbial genomes.

Roadmap for future research: systems biology and beyond

Holobiont paradigm and microbiome-driven resilience

One of the most forward-looking themes in this Research Topic is the plant holobiont concept—the recognition that plant performance is co-determined by its microbiome:

- Tyagi et al. reviewed microbial dysbiosis under waterlogging stress.
- Ahmad et al. showed how P. indica inoculation enhances salt tolerance in date palm.

These studies collectively advocate for "holobiont breeding", integrating host genetic traits with microbiome function—a paradigm shift from conventional plant-centric approaches to co-optimized plant–microbiome systems.

Towards systems-level crop improvement: challenges and horizons

The articles in this Research Topic signal a transformation in plant biology—from single-gene studies to systems-level, multidimensional investigations. Key future directions identified include:

- Causal microbiome engineering uses synthetic communities and functional metagenomics.
- Temporal and spatial gene expression resolution through high-resolution time-series and cell-type-specific profiling.
- Pan-genome analysis and structural variant discovery beyond SNP-centric studies.
- CRISPR/dCas9 epigenome editing and synthetic RNA technologies for precision trait modulation.
- AI-driven phenotype prediction and climate-resilient variety design through integration of omics and environmental data.

These trajectories call for transdisciplinary collaboration, uniting plant genetics, epigenomics, microbiology, data science, and field-based agronomy.

The 17 articles published under this Research Topic represent a significant leap forward in our collective understanding of plant stress biology. These studies span a diverse array of plant systems, stress types, and methodological frameworks ranging from GWAS, meta-QTL analyses, transcriptomics, metabolomics, and epigenetics to field validation and microbiome profiling. Together, this Research Topic moves beyond single-stressor frameworks by embracing integrative approaches that capture the interactions among plant genomes, epigenomes, and the associated microbial communities-the plant holobiont. Multiple studies have demonstrated that stress resilience is not solely encoded in the plant genome but is also influenced by dynamic plant–microbiome interactions. For instance, insights from Liu et al. Deng et al., Karapareddy et al. and Tyagi et al. emphasize the role of rhizosphere and endosphere microbial shifts under stress and highlight beneficial microbes as potential bioenhancers. Simultaneously, genetic and genomic studies by Vutla et al., Sahu et al., Suresh et al. and Chandana et al. offer precise markers and candidate genes for resistance breeding through QTL mapping and GWAS. Several contributions exemplify the shift toward actionable trait discovery (Kavai et al. and Wambi et al.) bridging lab-based genomic insights with field-based validation. Moreover, integrative molecular approaches including miRNA profiling (Patil and Tripathi), transcriptome analysis (Ahmad et al.), and epigenetic regulators (Su et al.) advance our mechanistic understanding of stress perception and response. The evolutionary insights from Sultan et al. add further depth by situating stress responses within the broader context of gene family diversification and adaptation. Collectively, these works signal a notable advancement toward designing crops not merely as standalone genetic entities, but as dynamic systems interacting continuously with their environment and associated biota. A compelling future direction emerging from these studies is the incorporation of microbiome-aware selection and breeding a concept gaining traction as "holobiont breeding" (Huitzil et al., 2023). This requires precise characterization of beneficial microbial consortia, their functions, and the plant traits that facilitate their recruitment and persistence under stress.

The future of plant resilience research must integrate multi-omics with high-throughput phenotyping, environmental modeling, and artificial intelligence. Such convergence will enable the discovery of novel gene networks, predictive trait–microbe associations, and adaptive alleles suited to future climates (Wang et al., 2024). Structural variants and presence/absence variations uncovered through pangenomics and Pan-GWAS offer another layer of stress-relevant variation often missed in SNP-only frameworks (Hu et al., 2024, 2025). Advances in genome editing tools like CRISPR/dCas systems and synthetic biology platforms hold promise for targeted modulation of regulatory networks involved in stress tolerance (Miki et al., 2018). Moreover, future research must prioritize stress combinations and field-level complexity. Multifactorial stress experiments, longitudinal sampling, cell-type specific profiling, and synthetic community reconstitution are necessary to elucidate causal interactions and validate lab findings under real-world agricultural scenarios (Schmitz et al., 2022; Taskiran et al., 2024).

Conclusion

The convergence of climate variability, pathogen pressure, abiotic stress, and soil degradation poses complex challenges to global agriculture. Addressing these requires an integrated understanding of plant genetics, physiology, and plant–microbe interactions. The Research Topic offers a blueprint for the future of sustainable crop improvement. It calls for systems-level integration of plant genetics, microbial ecology, functional genomics, and field-based validation. The holistic vision presented here anchored in robust data and interdisciplinary collaboration will be pivotal in breeding the next generation of climate-smart, microbiome-optimized, and climate-adapted varieties. We are confident that the insights presented in this Research Topic will inspire continued innovation and strategic research toward global food and environmental security. The studies presented herein converge on a singular message: the path to climate-smart agriculture lies in holobiont-aware breeding, where plant genotypes and their associated microbiomes co-evolve and co-adapt for sustainable productivity.

We extend our deepest appreciation to the authors, reviewers, and editorial teams whose contributions make this thematic issue a robust resource and a visionary roadmap for advancing plant resilience research. The insights and innovations presented here pave the way for a future where crops are not only genetically

fortified but ecologically and epigenetically empowered to thrive in the face of multifactorial stresses.

Author contributions

SP: Validation, Project administration, Writing – review & editing, Formal Analysis, Funding acquisition, Data curation, Investigation, Methodology, Conceptualization, Resources, Writing – original draft. MT: Validation, Formal Analysis, Data curation, Writing – review & editing, Conceptualization, Methodology, Writing – original draft, Resources, Visualization. RM: Resources, Writing – review & editing, Writing – original draft, Data curation, Validation, Visualization.

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 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.

The authors MT, RM 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

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

References

Fedoroff, N. V., Battisti, D. S., Beachy, R. N., Cooper, P. J., Fischhoff, D. A., Hodges, C. N., et al. (2010). Radically rethinking agriculture for the 21st century. *Science* 327, 833–834. doi: 10.1126/science.1186834

Hu, H., Wang, J., Nie, S., Zhao, J., Batley, J., and Edwards, D. (2024). Plant pangenomics, current practice and future direction. *Agric. Commun.* 2, 100039. doi: 10.1016/j.agrcom.2024.100039

Hu, H., Zhao, J., Thomas, W. J., Batley, J., and Edwards, D. (2025). The role of pangenomics in orphan crop improvement. *Nat. Commun.* 16, 118. doi: 10.1038/s41467-024-55260-4

Huitzil, S., Huepe, C., Aldana, M., and Frank, A. (2023). The missing link: how the holobiont concept provides a genetic framework for rapid evolution and the inheritance of acquired characteristics. *Front. Ecol. Evol.* 11. doi: 10.3389/fevo.2023.1279938

Jiang, Z., van Zanten, M., and Sasidharan, R. (2025). Mechanisms of plant acclimation to multiple abiotic stresses. *Commun. Biol.* 8, 655. doi: 10.1038/s42003-025-08077-w

Miki, D., Zhang, W., Zeng, W., Feng, Z., and Zhu, J.-K. (2018). CRISPR/Cas9-mediated gene targeting in *Arabidopsis* using sequential transformation. *Nat. Commun.* 9, 1967. doi: 10.1038/s41467-018-04416-0

Schmitz, L., Yan, Z., Schneijderberg, M., de Roij, M., Pijnenburg, R., Zheng, Q., et al. (2022). Synthetic bacterial community derived from a desert rhizosphere confers salt stress resilience to tomato in the presence of a soil microbiome. *ISME J.* 16, 1907–1920. doi: 10.1038/s41396-022-01238-3

Taskiran, I. I., Spanier, K. I., Dickmänken, H., Kempynck, N., Pančíková, A., Ekşi, E. C., et al. (2024). Cell-type-directed design of synthetic enhancers. *Nature* 626, 212–220. doi: 10.1038/s41586-023-06936-2

Wang, P., Lehti-Shiu, M. D., Lotreck, S., Segura Abá, K., Krysan, P. J., and Shiu, S. H. (2024). Prediction of plant complex traits via integration of multi-omics data. *Nat. Commun.* 15, 6856. doi: 10.1038/s41467-024-50701-6

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 28 May 2024
DOI 10.3389/fpls.2024.1381753

Check for updates

OPEN ACCESS

EDITED BY
Somashekhar M. Punnuri,
Fort Valley State University, United States

REVIEWED BY
Sunil K. Kenchanmane Raju,
Michigan State University, United States
Zihao Zheng,
Syngenta, United States

*CORRESPONDENCE
Xiaozhao Xu
201801006@qau.edu.cn
Chenxia Cheng
chengchenxia@qau.edu.cn

RECEIVED 22 February 2024
ACCEPTED 13 May 2024
PUBLISHED 28 May 2024

CITATION
Su S, Ji M, Chen J, Zhang M, Xu X and Cheng C (2024) Genome-wide identification and expression analysis of protein arginine methyltransferase and JmjC domain-containing family in apple.
Front. Plant Sci. 15:1381753.
doi: 10.3389/fpls.2024.1381753

COPYRIGHT
© 2024 Su, Ji, Chen, Zhang, Xu and Cheng. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Genome-wide identification and expression analysis of protein arginine methyltransferase and JmjC domain-containing family in apple

Shenghui Su[1], Min Ji[2], Jiaqi Chen[3], Meidie Zhang[4], Xiaozhao Xu[1,5,6]* and Chenxia Cheng[1]*

[1]College of Horticulture, Qingdao Agricultural University, Qingdao, Shandong, China, [2]Engineering Laboratory of Genetic Improvement of Horticultural Crops of Shandong Province, Qingdao Agricultural University, Qingdao, China, [3]National Technology Centre for Whole Process Quality Control of FSEN Horticultural Products (Qingdao), Qingdao Agricultural University, Qingdao, China, [4]Laboratory of Quality & Safety Risk Assessment for Fruit (Qingdao), Ministry of Agriculture and Rural Affairs, Qingdao Agricultural University, Qingdao, China, [5]Qingdao Key Lab of Modern Agricultural Quality and Safety Engineering, Qingdao Agricultural University, Qingdao, China, [6]Academy of Dongying Efficient Agricultural Technology and Industry on Saline and Alkaline Land in Collaboration with Qingdao Agricultural University, Dongying, China

Histone methylation is an important type of histone modification that regulates gene expression in plants. In this study, we identified 14 arginine methylation-related genes (*Protein Arginine Methyltransferase*, *MdPRMT*) and 32 demethylation-related genes (*JmjC Domain-Containing Family*, *MdJMJ*) in apple. Furthermore, we investigated the phylogenetic relationship, chromosome distribution, gene structure, motif analysis, promoter sequence analysis, and expression patterns of *MdPRMT* and *MdJMJ* genes. Homology analysis showed a high degree of conservation and homology between *PRMT* and *JMJ* genes in *Arabidopsis* and apple. We identified the types of duplicated genes in the *MdJMJ* and *MdPRMT* gene families, found a large number of whole-genome duplicates (WGD) gene pairs and a small number of tandem duplicates (TD) pairs, transposed duplication (TRD) gene pairs as well as proximal duplicates (PD) pairs, and discussed the possible evolutionary pathways of the gene families from the perspective of duplicated genes. Homology analysis showed a high degree of conservation and homology between *PRMT* and *JMJ* genes in *Arabidopsis* and apple. In addition, the promoter regions of *MdPRMT* and *MdJMJ* contain numerous *cis*-acting elements involved in plant growth and development, hormone response, and stress responses. Based on the transcriptional profiles of *MdPRMT* and *MdJMJ* in different tissues and developmental stages, it was found that *MdPRMT* and *MdJMJ* may play multiple roles in apple growth and development, for example, *MdJMJ21* may be involved in the regulation of apple endosperm formation. *MdPRMT* and *MdJMJ* exhibit different expression patterns

in response to hormone signaling in apple, *MdJMJ3*, *MdJMJ18*, *MdJMJ30*, *MdPRMT2*, *MdPRMT13*, and *MdPRMT14* may play roles in apple response to drought stress, while the expression of *MdJMJ13*, *MdPRMT3*, *MdPRMT4*, and *MdPRMT6* is affected by cold stress. Our study provides a foundation for determining the functional roles of *MdPRMT* and *MdJMJ* genes in apple.

KEYWORDS

Prmt, JMJ, apple, gene expression, gene family

1 Introduction

In eukaryotes, the basic structural unit of chromatin is the nucleosome (Kouzarides, 2007). Generally, nucleosomes are composed of octamers consisting of two copies each of four histone subunits, H2B, H2A, H3, and H4 (Thomas Jenuwein, 2001; Liu et al., 2010). The N-terminal amino acid residues of histones are susceptible to post-translational modifications, including acetylation, methylation, phosphorylation, ubiquitination, and other histone modifications (Kouzarides, 2007). In recent years, with the advancement in detection techniques for histone modifications and further research, it has been discovered that the middle and C-terminal regions of histones can also be specifically modified. These modifications affect the compactness and accessibility of chromatin in different ways, thereby influencing gene expression and ultimately affecting various physiological and developmental processes in organisms. They are one of the most important epigenetic regulatory mechanisms of gene expression in eukaryotes (Lawrence et al., 2016). Histone modifications are reversible covalent modifications. The occurrence, removal, and functional roles of these covalent modifications are mainly regulated by histone-modifying enzymes and their corresponding cofactors, including Writers, Erasers, and Reader/Effectors (Liu et al., 2010). Writers are enzymes that catalyze the addition of chemical groups to histones for modification, such as histone acetyltransferases (HATs), histone methyltransferases (HMTs), kinases, and ubiquitinases. Erasers are enzymes that remove these modifications from histones, such as histone deacetylases (HDACs), histone demethylases (HDMs), phosphatases, and deubiquitinases. Readers are proteins or protein complexes that recognize and specifically bind to substrates with specific post-translational modifications (Liu et al., 2010). Among them, histone methylation, as one of the main types of histone modifications, primarily affects the binding of histones to Reader proteins, leading to changes in chromatin structure, and thus transcriptional repression or activation (Genschik et al., 2014). Histone methylation can be divided into lysine methylation and arginine methylation based on their occurrence sites (Alvarez-Venegas, 2005), so HMTs can also be classified into HKMTs and PRMTs. The removal of histone methylation modifications is mainly accomplished by HDMs (Luo et al., 2014).

Protein arginine methylation is catalyzed by type I and II PRMTs, with type I PRMTs catalyzing asymmetric dimethylation of arginine residues and type II PRMTs catalyzing symmetric dimethylation of arginine residues (Bedford and Richard, 2005). Studies have shown that *Arabidopsis* protein arginine methyltransferase 5 (AtPRMT5), a homolog of human PRMT5, is an enzyme capable of catalyzing symmetric dimethylation of arginine residues, and it plays a role in *Arabidopsis* growth and development, especially in promoting flowering (Wang et al., 2001; Pei et al., 2007). In addition, other type I PRMTs in *Arabidopsis* that can catalyze asymmetric dimethylation of arginine residues (*AtPRMT10*, *AtPRMT4a*, and *AtPRMT4b*) have also been shown to promote flowering (Niu et al., 2007, 2008). Overall, existing evidence suggests that protein arginine methyltransferases play important roles in plant growth and development. HDMs are key factors regulating the steady state of histone methylation. HDMs can be divided into two types based on their mechanisms: lysine-specific demethylase KDM1/LSD1 and demethylases containing Jumonji C (JmjC) domains. KDM1/LSD1 performs demethylation of mono- and dimethylated lysine residues through FAD-dependent amine oxidation reactions, while JmjC proteins catalyze demethylation reactions that are dependent on iron (II) and α-ketoglutarate, and they play important roles in histone demethylation. They can remove methylation modifications at H3K4, H3K9, H3K27, and H3K36 (Anand and Marmorstein, 2007). *Arabidopsis* contains 21 JmjC domain proteins (JMJs), which are divided into five subfamilies (KDM5/JARID1, KDM4/JHDM3, KDM3/JHDM2, JMJD6, and JmjC domain-only) based on protein sequence similarity (Luo et al., 2014). In rice, 20 JmjC domain proteins have been identified, among which OsJMJ706 specific to H3K9 and OsJMJ703 specific to H3K4 have been shown to be involved in the regulation of rice flower organ development and transposon silencing, respectively (Sun and Zhou, 2008; Copenhaver et al., 2013; Cui et al., 2013). These findings collectively indicate that PRMTs and HDMs play important roles in plant growth and development.

Apple (*Malus domestica*) is one of the most important fruits in the world (Perini et al., 2014). Previous studies have shown that histone lysine methylation in apples is primarily mediated by a class of proteins containing SET domains (SDGs). The apple genome

contains a total of 67 *SDG* genes, which play important roles in apple development and stress responses (Li et al., 2021). However, research on protein arginine methyltransferases and histone demethylases in apples is very limited, so it is necessary to reveal their functional characteristics in apples.

In this study, a total of 32 JMJ members and 14 PRMT members were identified in the apple genome, and they were subjected to detailed analysis in terms of systematic evolution, homology relationships, conserved domains, gene structure, and *cis*-acting elements. In addition, the expression profiles of *MdPRMTs* and *MdMJMs* in different organs of apple at different developmental stages, under biotic and abiotic stresses, and in response to hormones were also studied. In conclusion, this study provides a comprehensive analysis of *JMJ* and *PRMT* genes in apple, laying a foundation for further exploration of their regulatory roles in apple development, stress responses, and hormone signaling.

2 Materials and methods

2.1 Identification and classification of histone methylation modification genes

Following previous research, the *Arabidopsis* Information Resource (TAIR, https://www.arabidopsis.org/) was used to obtain the *Arabidopsis thaliana* histone methylation modification genes protein sequences. The whole genome data of apples (*Malus domestica*) was obtained from the Apple Genome Database (http://bioinformatics.cau.edu.cn/AppleMDO/) (Da et al., 2019). And the gene identifier (ID) can be found in the Supplementary Table S1.

Candidate histone methylation modification genes (*PRMTs* and *JMJs*) in apple were identified using; two basic local alignment search tool (BLAST) methods implemented in TBtools (Chen et al., 2020). The *Arabidopsis* PRMTs and JMJs protein sequences were used as queries for BLAST analysis. Additionally, relevant hidden Markov models were obtained from the Pfam database (http://pfam.xfam.org/), and employed to search the apple protein sequence data using the HMMER 3.0 software with a stringent threshold of $E \leq 10^{-20}$. After removing redundancy and duplicate sequences, a preliminary set of candidate sequences was determined. The intersection of the gene family candidate members was obtained by combining the results obtained from the above-mentioned methods. Subsequently, the conserved protein domains of the target gene family in apples were analyzed using the Web CD-Search tool on the NCBI website (https://www.ncbi.nlm.nih.gov/Structure/bwrb/bwrpsb.cgi/). This analysis aimed to determine whether the conserved domain related to the target gene family protein was present in each candidate sequence. Only the candidate sequences containing complete domains were retained for further analysis. In the case of those sequences that contained different or incomplete domains, their sequence integrity was assessed by submitting them to the SoftBerry website (http://linux1.softberry.com/) before subjecting them to the Batch Web CD-Search tool for comparing the similarity between their domains and those of *Arabidopsis* PRMTs and JMJs sequences.

The protein physicochemical properties of gene family members were calculated using the ExPASy online software ProtParam (https://web.expasy.org/protparam/). These properties encompass various parameters such as the number of amino acids, molecular weight, theoretical pI, instability index, aliphatic index, and grand average of hydropathicity. The chromosome location of each confirmed histone methylation modification gene was retrieved from the GFF3 file of apple genome, and then visualized on apple chromosomes using TBtools (Chen et al., 2020).

2.2 Phylogenetic analysis of histone methylation modification proteins in *M. domestica* and *A. thaliana*

The amino acid sequences of PRMTs, and JMJs in *Arabidopsis* and apple were obtained individually from the TAIR and AppleMDO databases. These sequences were aligned using MUSCLE in MEGA 11 software. Then, the optimal model was determined using MEGA 11 software, and the phylogenetic tree was constructed using the maximum likelihood method(ML) (Tamura et al., 2021). The bootstrap consensus tree inferred from 1000 replicates is taken to represent the evolutionary history of the taxa analyzed (Felsenstein, 1985). The phylogenetic tree was visualized and optimized by TBtools.

2.3 Chromosomal distribution and synteny analysis

The gene location analysis was performed using the Amazing Gene Location tool in TBtools, which mapped the *PRMTs* and *JMJs* to their respective chromosomes based on the information extracted from the GTF/GFF files. To visualize the synteny blocks within the apple genome and between the genomes of apple, *Arabidopsis*, rice and maize, the Dual Synteny Plotter tool in TBtools was employed. These synteny blocks were generated using the One Step MCScanX-Super Fast method. Rice and maize genome data were both downloaded from the EnsemblPlants database (https://plants.ensembl.org/index.html). The Simple Ka/Ks calculator (NG) in TBtools was utilized to calculate the synonymous (Ks) and non-synonymous (Ka) values for duplicated genes.

2.4 Identification of duplicated gene types

We employed the DupGen_finder tool (https://github.com/qiao-xin/DupGen_finder) to classify the types of duplicated genes. DupGen_finder_unique categorizes plant genome duplicated genes into five categories based on a specific algorithm: whole-genome duplication (WGD), proximal duplication (PD), transposed duplication (TRD), dispersed duplication (DSD), and tandem

duplication (TD) (Qiao et al., 2019). Using *Arabidopsis* as an outgroup for apple, we followed the recommended workflow of DupGen_finder for our analysis and parameter configuration as previously described (Kenchanmane Raju et al., 2023).

2.5 Sequence analysis

The conserved features of PRMTs and JMJs sequences were analyzed and visualized based on motifs using the Multiple Em for Motif Elicitation (MEME) suite 5.4.1 (https://meme-suite.org/meme/tools/meme) (Bailey et al., 2015) and TBtools. The gene structure of PRMT and JMJ genes were assessed using Gene Structure Shower in TBtools, utilizing information from the apple genome GFF3 file. To identify the cis-elements in the promoters of *PRMT* and *JMJ* genes, PlantCare (https://bioinformatics.psb.ugent.be/webtools/plantcare/html/) was employed (Lescot et al., 2002).

2.6 Expression profiling of *PRMT* and *JMJ* genes in apple

The expression pattern of *PRMT* and *JMJ* genes following abiotic, hormone and biotic stress treatment were obtained from SRA database in NCBI website or National Genomics Data Center. The study IDs were as follows: for abiotic stress treatment: salt treatment (SRP229388), cold stress (CRA002596) and drought treatment (SRP347250); for hormone treatment: 1-Methylcyclopropene (1-MCP) treatment (SRP334206), Gibberellin Acid 3 (GA_3) and 1-Naphthaleneacetic acid (NAA) treatment (SRP185711) and Indole-3-butyric acid (IBA) treatment (SRP330812); for biotic stress treatment: *Alternaria alternata* (SRP091754), *Pythium ultimum* (SRP048684), *Fusarium solani* (SRP239526) and ring rot (SRP153065). Kallisto, an RNA-seq quantification program were used to calculate the expression of gene (Bray et al., 2016). In this study, the gene expression was estimated using the fragment number per kilobase per million mapped exons (FPKM) and transcripts per kilobase of exon model per million mapped reads (TPM) method.

2.7 Quantitative reverse transcription polymerase chain reaction

Apple ('GL3') tissue culture seedlings were transferred to MS medium plus 200 mM sodium chloride (NaCl), 10% polyethylene glycol (PEG) 6000. Leaves were sampled at 0, 6, 12 and 24h after treatment for RNA isolation. Total RNA was extracted from apple leaves using the hot borate method as previously described (Ma et al., 2006). For reverse transcription-PCR, first-strand cDNA was synthesized from 1 μg of total RNA using Reverse Transcriptase M-MLV (RNase H-) (Takara Biomedical Technology Co., Ltd., Shiga, Japan). qRT-PCR was conducted with Takara SYBR Premix Ex Taq II (Takara Biomedical Technology Co., Ltd.) using a Light Cycler 480 instrument (Roche, Basel, Switzerland). *MdEF-1α* were used as internal controls. All primers used are listed in Supplementary Table S8.

TABLE 1 *PRMT and JMJ* genes identified in apple.

Gene family	Gene name	Gene ID	Chro mosome	Genomic location	GRAVY	Strand	Length (aa)	pI	MW (kDa)
PRMTs	*MdPRMT1*	MD02G1037100	chr2	2,800,117–2,803,571	-0.455	+	383	5.43	43.09
	MdPRMT2	MD02G1074300	chr2	5,963,796–5,966,843	-0.261	+	393	5.11	44.16
	MdPRMT3	MD04G1063800	chr4	8,432,676–8,437,714	-0.216	+	722	5.99	81.12
	MdPRMT4	MD07G1046100	chr7	3,921,653–3,932,752	-0.325	–	636	4.6	70.02
	MdPRMT5	MD08G1132700	chr8	12,539,643–12,547,230	-0.259	+	649	5.55	72.7
	MdPRMT6	MD13G1002200	chr13	136,755–142,344	-0.4	–	423	5.9	47.63
	MdPRMT7	MD13G1168500	chr13	13,630,857–13,636,161	-0.284	–	545	5.31	60.72
	MdPRMT8	MD15G1111800	chr15	7,855,147–7,857,854	-0.098	–	241	5.33	27.34
	MdPRMT9	MD15G1112100	chr15	7,859,494–7,866,979	-0.244	–	649	5.62	72.56
	MdPRMT10	MD15G1112600	chr15	7,919,669–7,923,893	-0.189	–	421	5.87	47.34
	MdPRMT11	MD15G1112700	chr15	7,923,895–7,927,030	-0.077	–	232	4.61	25.45
	MdPRMT12	MD15G1177300	chr15	13,877,192–13,880,754	-0.381	+	382	5.28	42.85
	MdPRMT13	MD15G1203800	chr15	16,210,542–16,213,676	-0.266	+	391	5.06	43.91
	MdPRMT14	MD16G1168900	chr16	14,047,574–14,053,079	-0.228	–	552	5.33	61.36
JMJs	*MdJMJ1*	MD00G1097500	chr0	20,399,985–20,405,925	-0.548	–	1042	5.63	117.30
	MdJMJ2	MD00G1097600	chr0	20,412,328–20,413,454	-0.473	–	187	5.73	21.43
	MdJMJ3	MD01G1082300	chr1	18,879,413–18,889,690	-0.921	–	1099	8.51	126.12

(Continued)

10.3389/fpls.2024.1381753

TABLE 1 Continued

Gene family	Gene name	Gene ID	Chro mosome	Genomic location	GRAVY	Strand	Length (aa)	pI	MW (kDa)
	MdJMJ4	MD01G1103000	chr1	21,501,838–21,506,502	-0.213	–	903	8.97	97.96
	MdJMJ5	MD01G1106000	chr1	21,918,410–21,925,954	-0.55	–	1603	5.67	176.95
	MdJMJ6	MD01G1218500	chr1	31,068,432–31,075,270	-0.681	+	1519	8.34	166.66
	MdJMJ7	MD03G1220300	chr3	30,449,337–30,452,592	-0.195	–	748	6.6	82.31
	MdJMJ8	MD04G1202800	chr4	28,862,118–28,866,548	-0.457	+	600	5.05	67.77
	MdJMJ9	MD04G1229800	chr4	30,997,445–31,004,908	-0.591	–	1215	6.36	135.97
	MdJMJ10	MD05G1326700	chr5	45,286,027–45,292,287	-0.562	–	1024	8.43	116.05
	MdJMJ11	MD05G1351300	chr5	46,832,063–46,837,369	-0.677	–	887	6.07	101.73
	MdJMJ12	MD06G1012500	chr6	1,590,799–1,599,810	-0.869	–	1046	7.31	118.43
	MdJMJ13	MD06G1026100	chr6	3,210,489–3,217,400	-0.509	+	890	7.09	99.49
	MdJMJ14	MD06G1081900	chr6	20,068,477–20,088,135	-0.316	–	1843	6.52	210.00
	MdJMJ15	MD06G1159300	chr6	30,080,831–30,088,073	-0.719	–	1467	8.87	164.00
	MdJMJ16	MD07G1099600	chr7	10,973,691–10,988,237	-0.829	–	1233	5.98	139.42
	MdJMJ17	MD07G1172400	chr7	24,961,209–24,968,468	-0.535	–	1594	5.84	176.35
	MdJMJ18	MD08G1004400	chr8	412,295–414,742	-0.204	+	815	5.28	89.09
	MdJMJ19	MD08G1186800	chr8	23,531,368–23,540,505	-0.317	+	981	5.32	111.64
	MdJMJ20	MD10G1182700	chr10	27,556,246–27,561,506	-0.568	+	885	7.09	98.57
	MdJMJ21	MD10G1241100	chr10	33,658,731–33,664,694	-0.537	–	1040	5.53	117.64
	MdJMJ22	MD10G1304800	chr10	39,112,395–39,118,408	-0.604	–	1030	8.57	116.82
	MdJMJ23	MD10G1325700	chr10	40,607,236–40,611,894	-0.578	–	689	5.8	79.21
	MdJMJ24	MD12G1046300	chr12	5,202,290–5,205,217	-0.357	–	516	6.78	58.43
	MdJMJ25	MD12G1216600	chr12	29,418,485–29,426,130	-0.636	+	946	5.75	107.82
	MdJMJ26	MD12G1246900	chr12	31,721,067–31,728,936	-0.593	–	1236	7.14	138.65
	MdJMJ27	MD14G1103700	chr14	15,999,174–16,014,561	-0.304	–	1842	7.67	209.35
	MdJMJ28	MD14G1165600	chr14	25,922,669–25,930,130	-0.693	–	1466	9.03	163.77
	MdJMJ29	MD14G1175900	chr14	26,919,946–26,923,646	-0.416	+	474	5.32	55.23
	MdJMJ30	MD15G1003800	chr15	233,872–236,319	-0.209	+	815	5.14	88.85
	MdJMJ31	MD15G1372700	chr15	45,518,685–45,527,976	-0.287	+	975	5.29	110.42
	MdJMJ32	MD16G1280000	chr16	38,164,869–38,171,753	-0.454	–	888	6.91	99.22

3 Results

3.1 Identification of *PRMTs* and *JMJs* in apple

To identify the *PRMTs* and *JMJs* gene families in the apple genome, a comprehensive investigation was carried out using the protein sequences of the corresponding genes from *Arabidopsis* as search queries. After eliminating redundant and duplicate sequences, we identified 14 and 32 genes in PRMT and JMJ families, respectively, in the apple genome (Table 1). The predicted protein sequence lengths of the PRMT and JMJ families ranged from 232 (MdPRMT11) to 722 (MdPRMT3) and 187 (MdJMJ2) to 1843 (MdJMJ14) amino acids, respectively. The molecular weight (MW) of the PRMT and KDM families ranged from 25.45 (MdPRMT11) to 81.12 (MdPRMT3) and 21.43 (MdJMJ2) to 210 (MdJMJ14) kDa, respectively. The isoelectric point (pI) values for the PRMT and JMJ families varied from 25.45 (MdPRMT11) to 81.12 (MdPRMT3) and 5.05 (MdJMJ8) to 9.03 (MdJMJ28), respectively. The complete information about the genes, including gene locus ID, chromosomal position, coding sequence length, and protein sequence length, was shown in Table 1. The nomenclature of the genes followed a sequential order based on their arrangement on the chromosome.

10.3389/fpls.2024.1381753

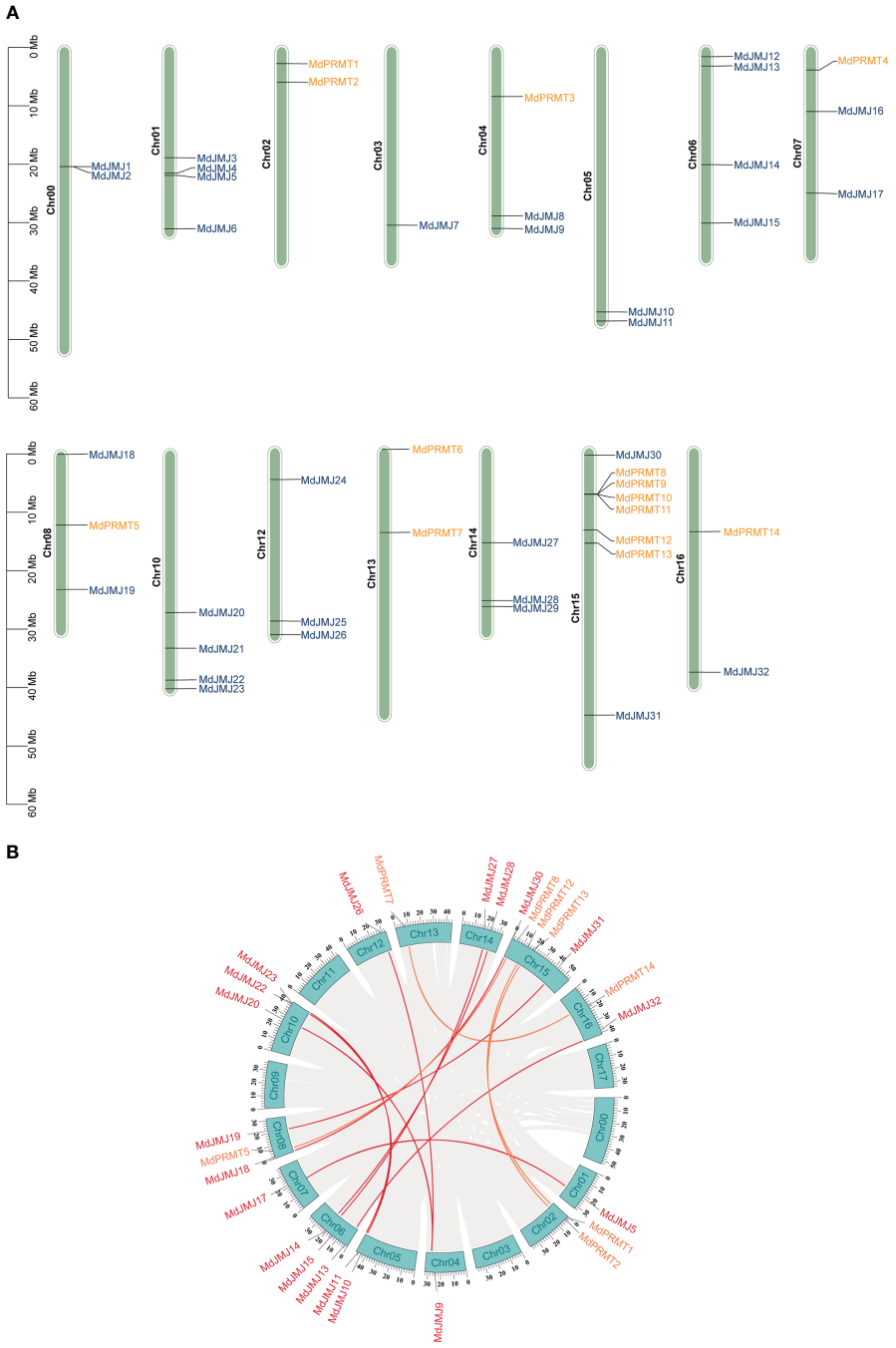


FIGURE 1
Chromosomal distribution and collinearity analysis of *JMJ* and *PRMT* genes between apple and *A. thaliana*. **(A)** Chromosomal distribution and duplication analysis of *JMJs* and *PRMTs* in apple. **(B)** Collinearity analysis of *JMJ* and *PRMT* genes family in apple.

3.2 Analysis of chromosomal distribution and evolutionary relationships of *MdJMJ* and *MdPRMT* genes

In Figure 1A, the chromosomal distribution of *PRMTs*, and *JMJs* in apple was shown. The analysis revealed that the distribution of these genes is irregular. Particularly, the *MdJMJ* genes were distributed on all chromosomes apart from chromosomes 2, 9, 11, and 17. The chromosomal distribution of *MdPRMT* genes was comparatively more restricted, limited to chromosomes 2, 4, 7, 8, 13, 15, and 16. Notably, chromosome 15 showed the most concentrated distribution with a total of 6 genes (*MdPRMT8*, *MdPRMT9*, *MdPRMT10*, *MdPRMT11*, *MdPRMT12* and *MdPRMT13*).

Gene duplication events are a common occurrence across all species, playing a crucial role in generating novel functional genes and driving species evolution (Xu et al., 2012). In order to explore gene duplication events within the *JMJ* and *PRMT* gene families in apple, we conducted genome synteny analysis. The analysis revealed

the presence of 10 potential duplicated genes in the *MdJMJ* gene family and 4 in the *MdPRMT* gene family (Figure 1B). This finding suggests that the expansion of the *MdJMJ* gene family is primarily driven by segmental duplication, as indicated by the high number of potential duplicated genes. Notably, there is a significantly gene cluster located on chromosome 15 for the *MdPRMT* gene family (Figure 1A).

Using *Arabidopsis* as an outgroup for apple, we followed the recommended workflow of DupGen_finder and found that after the divergence of the two species, the *MdJMJ* family contained 10 WGD-pairs, 1 TD-pairs (*MdJMJ1-MdJMJ2*), and 3 TRD-pairs (*MdJMJ3-MdJMJ6*, *MdJMJ4-MdJMJ7*, *MdJMJ11-MdJMJ12*). No DSD-pairs and PD-pairs were found in the *MdJMJ* family. The *MdPRMT* family had 4 WGD-pairs (*MdPRMT7- MdPRMT14*, *MdPRMT5-MdPRMT8*, *MdPRMT2-MdPRMT13*, *MdPRMT1-MdPRMT12*), 2 PD-pairs (*MdPRMT11-MdPRMT9*, *MdPRMT9-MdPRMT10*) and 1 TRD- pairs (*MdPRMT6-MdPRMT13*). No DSD-pairs and TD-pairs were found in the *MdPRMT* family. Detailed analytical results can be found in Supplementary Table S9. The analysis outcomes indicate that the primary mode of evolution for the *MdJMJ* and *MdPRMT* gene families is whole-genome duplication.

To elucidate the evolutionary relationships of the *PRMT* and *JMJ* gene families, a maximum likelihood (ML) phylogenetic tree was constructed using the PRMT and JMJ proteins from apple and *Arabidopsis* (Figure 2). The classification of the gene families was based on the established *Arabidopsis* JMJ gene families. The JMJ subfamily includes KDM5, KDM4, KDM3, KDM1, JMJC, and JMJD6, with 8, 6, 10, 4, 1, and 3 MdJMJ proteins, respectively (Figure 2A).

To gain deeper insights into the origin, evolutionary history, and functional characteristics of *PRMT* and *JMJ* genes, we performed a comparative synteny analysis between the genomes of apple, *Arabidopsis*, rice and maize (Figure 3). The apple genome contained 19 *JMJ* genes with high synteny to the *Arabidopsis*

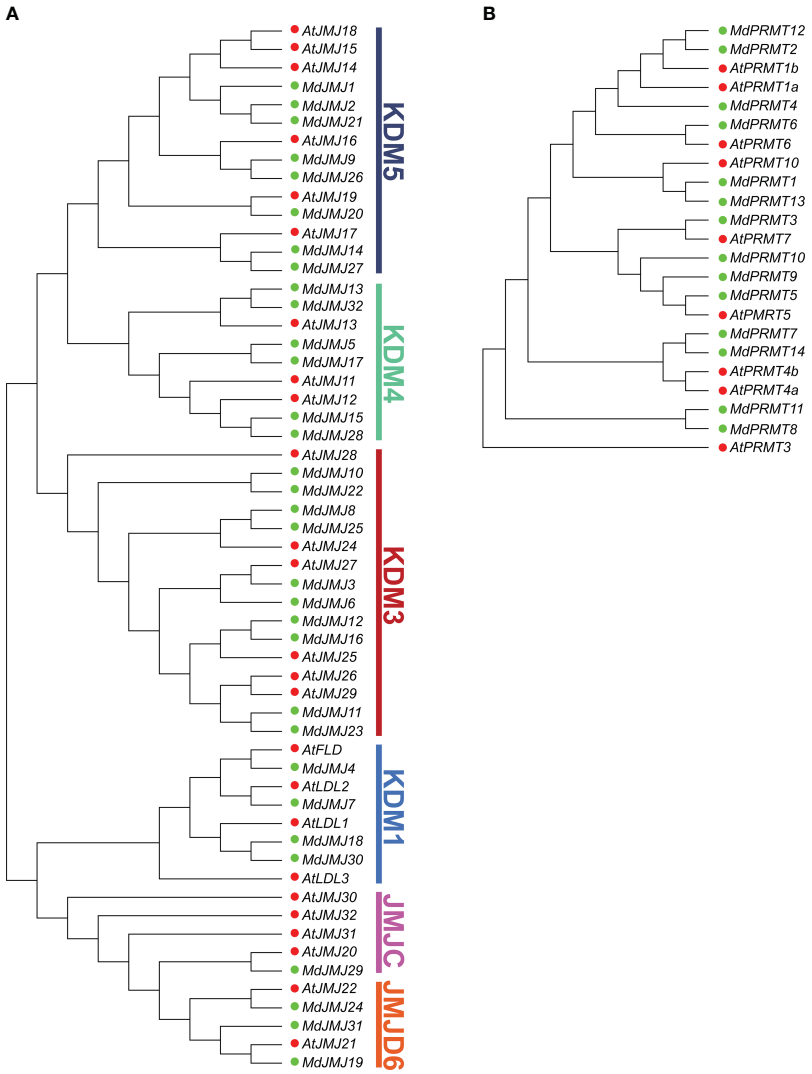


FIGURE 2
Phylogenetic relationship of JMJ **(A)** and PRMT **(B)** proteins from apple and *Arabidopsis*. The bootstrap consensus of tree was inferred from 1000 replicates, and the phylogenetic tree was constructed using the maximum likelihood method. In the figure, red dots represent *Arabidopsis* proteins, while green dots represent apple proteins.

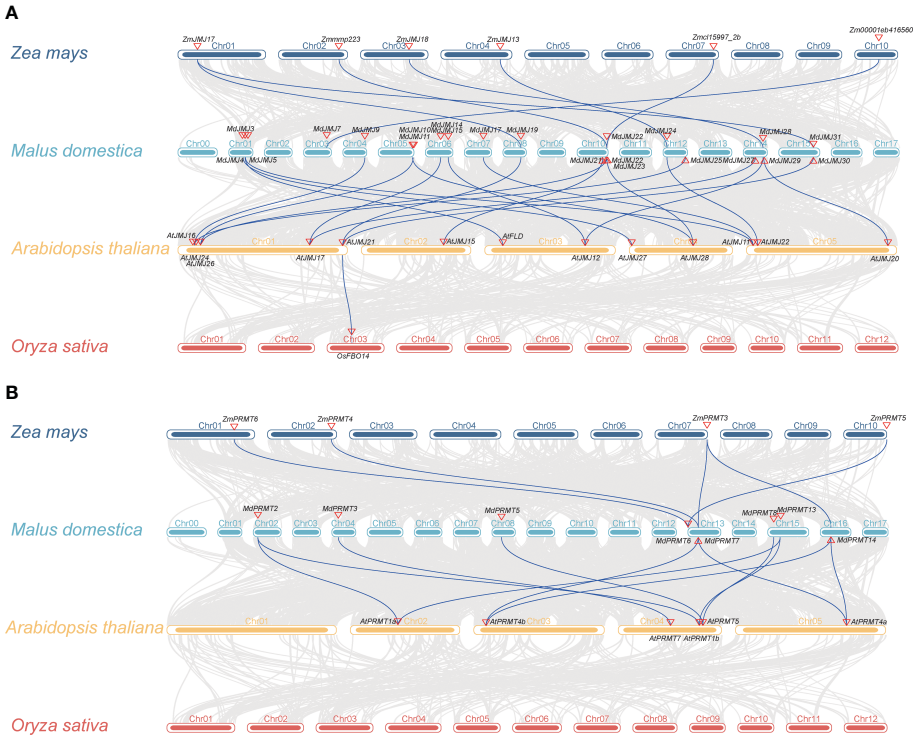


FIGURE 3
Synteny analysis of *JMJs* **(A)** and *PRMTs* **(B)** of apple, *Arabidopsis*, maize and rice.

genome, forming a total of 21 syntenic gene pairs, as shown in Figure 3A. These syntenic gene pairs include *MdJMJ10-AtJMJ16*, *MdJMJ11-AtJMJ17*, *MdJMJ11-AtJMJ18*, *MdJMJ14-AtJMJ19*, *MdJMJ15-AtJMJ20*, *MdJMJ17-AtJMJ21*, *MdJMJ19-AtJMJ22*, *MdJMJ21-AtJMJ23*, *MdJMJ22-AtJMJ24*, *MdJMJ23-AtJMJ25*, *MdJMJ23-AtJMJ26*, *MdJMJ24-AtJMJ27*, *MdJMJ25-AtJMJ28*, *MdJMJ27-AtJMJ29*, *MdJMJ28-AtJMJ30*, *MdJMJ29-AtJMJ31*, *MdJMJ3-AtJMJ32*, *MdJMJ31-AtJMJ33*, *MdJMJ4-AtJMJ34*, *MdJMJ5-AtJMJ35*, and *MdJMJ9-AtJMJ36* (Figure 3A). 7 syntenic gene pairs were found between apples and maize, only one ortholog pair (*AtJMJ21-OsFBO14*) were found in *Arabidopsis* and rice. In addition, there were 7 *PRMT* genes in the apple genome that showed synteny with the *Arabidopsis* genome, forming a total of 11 syntenic gene pairs. These syntenic gene pairs include *MdPRMT13-AtPRMT1a*, *MdPRMT13-AtPRMT1b*, *MdPRMT14-AtPRMT4b*, *MdPRMT14-AtPRMT4a*, *MdPRMT2-AtPRMT1a*, *MdPRMT2-AtPRMT1b*, *MdPRMT3-AtPRMT7*, *MdPRMT5-AtPRMT5*, *MdPRMT7-AtPRMT4b*, *MdPRMT7-AtPRMT4a*, and *MdPRMT8-AtPRMT5* (Figure 3B). Five pairs of co-linear genes (*ZmPRMT6-MdPRMT6*, *ZmPRMT3-MdPRMT7*, *ZmPRMT3-MdPRMT14*, *ZmPRMT4-MdPRMT7*, and *ZmPRMT5-MdPRMT6*) were identified between maize and apple, while no co-linear genes were found between *Arabidopsis* and rice. These results indicate a high conservation of *JMJ* and *PRMT* genes between the apple and *Arabidopsis* genomes.

For evolutionary studies, Ka (non-synonymous substitution) and Ks (synonymous substitution) values can be used to predict the selective pressure on duplicated genes (Zhao et al., 2016). When the Ka/Ks ratio falls below 1, it signifies that the genes have undergone purifying selection, indicating the absence of significant functional differentiation among these genes during the evolutionary process (Hurst, 2002). We utilized the Simple Ka/Ks CalculatorNG program in TBtools to analyze 14 pairs of collinear genes in *MdJMJ* and *MdPRMT* to further investigate the evolutionary relationships among these collinear genes. The results showed that all the collinear gene pairs had a Ka/Ks ratio less than 1, indicating that these genes have undergone purifying selection and have not experienced functional differentiation (Table 2). However, it is important to note that a Ka/Ks ratio less than 1 does not imply that these collinear genes have not undergone functional differentiation or evolution. It simply suggests that, within the observed time scale, the functional differences among these genes have had a relatively minor impact on the evolutionary process.

3.3 Sequence and structure analysis of JMJs and PRMTs

A phylogenetic tree for the MdJMJ and MdPRMT gene families was constructed using the Maximum Likelihood method in MEGA11.0 software (Figures 4A, E). It is evident that the genes clustered together in both gene families exhibit similar conserved domains. MdJMJ4 and MdJMJ31 contained only motif 1, while MdJMJ17 lacked any motifs. Additionally, motif 15, motif 19, and motif 20 were unique to MdJMJ19, MdJMJ29, MdJMJ28, and MdJMJ15, respectively (Figures 4B). The motif analysis was further

TABLE 2 Calculation of synonymous (Ks) and non-synonymous (Ka) for duplicated genes in apple.

Gene 1	Gene 2	Ka	Ks	Ka/Ks
MdJMJ5	*MdJMJ17*	0.06	0.17	0.35
MdJMJ9	*MdJMJ20*	0.44	1.84	0.24
MdJMJ9	*MdJMJ26*	0.05	0.16	0.29
MdJMJ10	*MdJMJ22*	0.04	0.15	0.28
MdJMJ11	*MdJMJ23*	0.05	0.16	0.29
MdJMJ15	*MdJMJ28*	0.04	0.14	0.29
MdJMJ14	*MdJMJ27*	0.05	0.15	0.32
MdJMJ13	*MdJMJ32*	0.06	0.14	0.42
MdJMJ19	*MdJMJ31*	0.04	0.20	0.22
MdJMJ18	*MdJMJ30*	0.03	0.15	0.18
MdPRMT2	*MdPRMT13*	0.02	0.15	0.12
MdPRMT1	*MdPRMT12*	0.02	0.18	0.10
MdPRMT5	*MdPRMT8*	0.10	0.26	0.37
MdPRMT7	*MdPRMT14*	0.02	0.16	0.14

supported by conserved domain analysis further supported the similar domain structures within gene cluster, which showed MdJMJ19, MdJMJ29, MdJMJ28, and MdJMJ15 possessed the ZZ superfamily domain, PLN02529 superfamily domain, and PLN02328 superfamily domain, respectively (Figures 4C). These three conserved domains were specific to these four genes, consistent with the motif analysis. Gene structure analysis revealed significant variation in the structure of *MdJMJ* genes, with clustering genes showing similar structures (Figure 4D). The number of introns ranged from 0 to 32, with *MdJMJ28* and *MdJMJ15* lacking introns, while *MdJMJ3* and *MdJMJ16* had the highest number of 32 introns.

For the MdPRMT gene family, which had a smaller number of members, clustering of genes based on conserved motifs revealed similar motif structures within the clusters (Figures 4F). Conserved domain analysis clearly reveals that MdPRMT10, MdPRMT5, MdPRMT2, and MdPRMT13 genes only contain the AdoMet_MTases conserved domain (Figures 4G). MdPRMT4 and MdPRMT8 genes only possess the COG4076 superfamily conserved domain. The PRMT5_C and PRMT5 conserved domains are present only in MdPRMT7, MdPRMT12, and MdPRMT1 genes. All the members of the *MdPRMT* gene family have gene structures with introns, ranging from 6 (*MdPRMT10* and

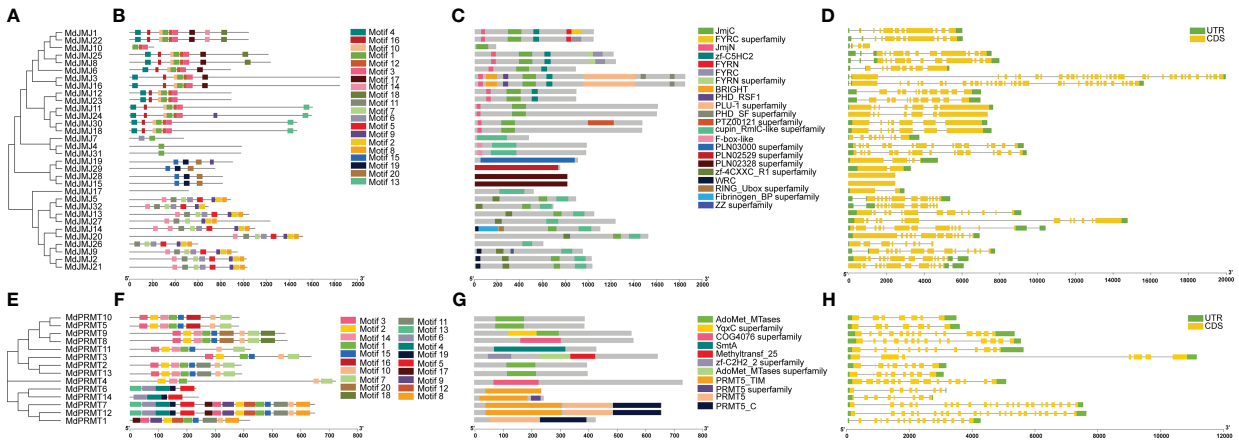


FIGURE 4
Phylogenetic analysis, motif composition, conserved domains of MdJMJ, and MdPRMT proteins and exon-intron structures of *MdJMJ* and *MdPRMT* genes. **(A–C)** The phylogenetic tree **(A)**, motif composition **(B)** and the distribution of conserved domains **(C)** of MdJMJs. **(D)** Gene structures of the *MdJMJ* genes. **(E–G)** The phylogenetic tree **(E)**, motif composition **(F)** and the distribution of conserved domains **(G)** of MdPRMTs. **(H)** Gene structures of the *MdPRMT* genes. Green rectangles represent untranslated regions (UTRs); yellow rectangles represent coding sequence (CDS) or exons; grey lines represent introns.

10.3389/fpls.2024.1381753

MdPRMT5) to 22 (*MdPRMT7* and *MdPRMT12*), indicating variations in gene structure within this gene family (Figures 4H).

The specific amino acid sequences of motifs of MdJMJ, MdPRMT and the detailed information of the MEME sites analysis were shown in the Supplementary Tables S2 and S3, respectively.

3.4 *Cis*-elements in the promoter of *MdPRMTs* and *MdJMJs*

An analysis was conducted on the promoter fragments (-2000 bp) of all *MdPRMT* and *MdJMJ* genes (Figures 5A, C). A large number of *cis*-regulatory elements were identified in their promoter regions, mainly including light response, cell cycle, low temperature, circadian rhythm, stress defense, growth and development (endosperm, seed-specific, meristem, cell differentiation), hormone signaling (MeJA, ABA, GA, Auxin, and SA), anoxic specific inducibility, and some transcription factors (TF), including MYB and ATBP. Additionally, we also found a wound responsive element in each of the *MdJMJ*. By counting the number of *cis*-regulatory elements (Figures 5B, D), it was found that light-responsive *cis*-elements were present in the promoter regions of all members of the *MdPRMT* and *MdJMJ*. MeJA, ABRE, and ARE elements were also widely distributed. Therefore, it can be inferred that *MdPRMTs* and *MdJMJs* play important roles in apple growth and development as well as hormone responses. On the other hand, less abundant *cis*-regulatory elements should not be overlooked. For example, the low-temperature elements were present in a significant number of *MdJMJs* and *MdPRMTs*, with 18 and 6 genes involved, respectively. Stress defense elements were found in 11, and 3 genes, while MYB-drought elements were present in 20 and 8 genes of *MdJMJ* and *MdPRMT*, respectively. These genes may play crucial roles in stress tolerance in apple. Additionally, one gene in *MdJMJs* contained an Endosperm-specific negative expression element and a Wound responsive element, while none of the *MdPRMTs* genes contained these two elements.

Similarly, to compare the differences in *cis*-regulatory elements between apple and *Arabidopsis*, the same method was used to construct a map of *cis*-regulatory elements in *Arabidopsis* promoters and categorize them according to the phylogenetic tree (Supplementary Figure S2). The results indicate that, on average, *Arabidopsis* has a fewer variety of *cis*-regulatory elements compared to apple. Like apple, *Arabidopsis AtJMJ* and *AtPRMT* genes also contain a large number of light-responsive and stress-response elements, which can provide some assistance for the function and evolution of the apple gene families.

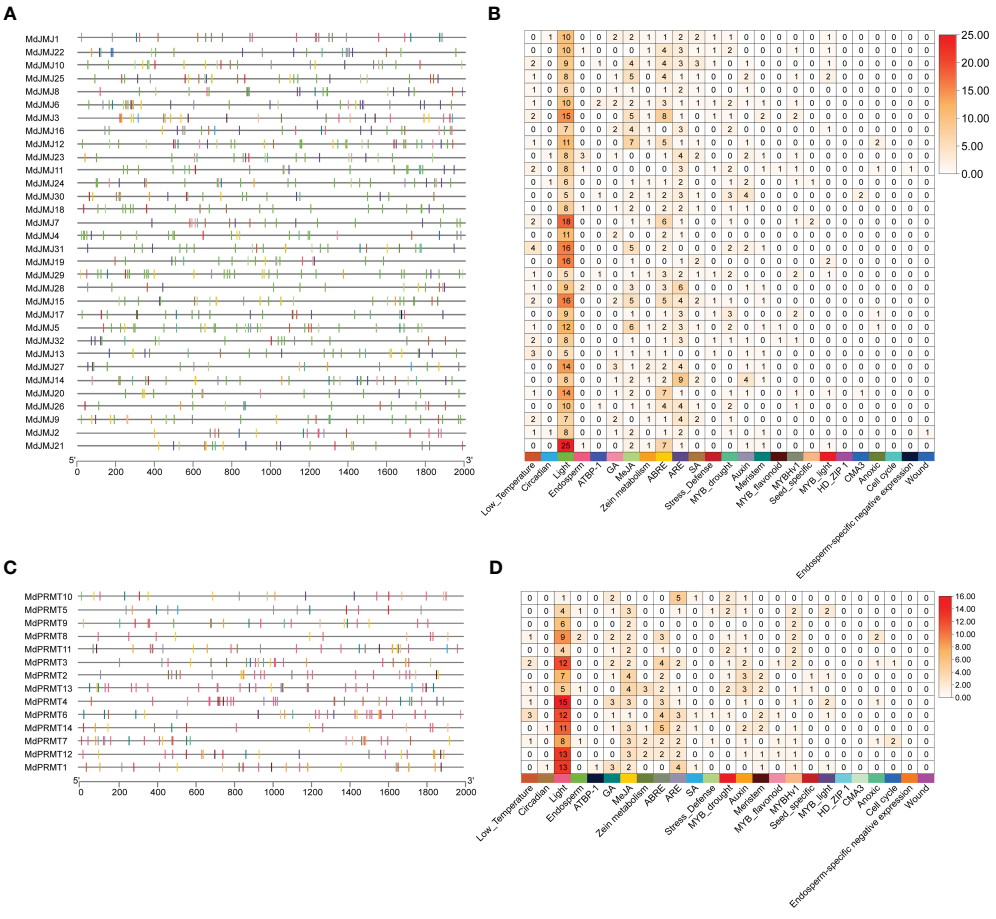


FIGURE 5
The *cis*-elements analysis of *MdJMJ* and *MdPRMT* promoters. **(A, B)** The distribution **(A)** and number **(B)** of *cis*-elements in the promoter of each *MdJMJ* genes. **(C, D)** The distribution **(C)** and number **(D)** of cis-elements in the promoter of each *MdPRMT* genes.

3.5 Expression profiles of the apple *MdPRMT* and *MdJMJ* in various organs, tissues and developmental stage

In this study, the expression profiles of *MdJMJs* and *MdPRMTs* in different tissues and developmental stages of apple were analyzed using publicly available transcriptome data (Figure 6). According to the results, the majority of histone modification-related genes showed relatively low expression levels during fruit maturation (SRP034165). It is worth noting that 11 histone modification-related genes (*MdPRMT1, MdPRMT2, MdPRMT4*, *MdPRMT5*, *MdPRMT6*, *MdPRMT7*, *MdPRMT12*, *MdPRMT13*, *MdPRMT14*, *MdJMJ6*, *and MdJMJ19*) exhibited higher expression levels in the early stage of fruit maturation, and their expression levels decreased to varying degrees as the fruit matured. Among them, the *MdPRMT* gene family had the largest number of genes, with approximately 64% of the genes showing decreased expression levels during fruit maturation. To delineate the expression patterns of histone modification-related genes in major organs of apple, transcriptome data from seedling and tree shoot apex, dormant buds and bud break, and different parts of apple flowers were analyzed. In shoot apex tissues at different stages, members of each gene family exhibited different expression patterns (SRP050139). For example, most members of the *MdPRMT* gene family showed varying degrees of down-regulation during the transition from the seedling stage to the mature stage, which was consistent with the down-regulated genes during fruit maturation. This suggests that the *MdPRMTs* may play an important role in the growth and development of apple. Furthermore, during apple plant development, 6 down-regulated genes were identified in the *MdJMJs*. Thus, it can be inferred that they collectively play a role in apple plant development.

Dormant buds play a crucial role in the growth cycle of plants and serve as a survival strategy for apple trees under adverse conditions such as winter or drought (Rohde and Bhalerao, 2007). They are also an important means of reproduction for apple trees. We analyzed the expression patterns of histone modification-related genes in apple dormant buds during dormancy and bud break (SPR099578). The results showed that *MdPRMTs* and *MdJMJs* had 5 and 6 genes, respectively, with higher expression levels. Among them, *MdPRMT7*, *MdJMJ3* and *MdJMJ25* showed relatively minor down-regulation, while the majority of genes exhibited lower expression levels or irregular trends.

Flowers are important reproductive organs of apple. We analyzed the transcriptome data from different parts of apple flowers. The results showed that histone modification-related genes had significantly higher average expression levels in pollen compared to other parts. *MdPRMTs*, and *MdJMJs* had 1 (*MdPRMT8*) and 5 (*MdJMJ3*, *MdJMJ13*, *MdJMJ21*, *MdJMJ23*, *MdJMJ32*) genes with significantly higher expression levels in pollen, respectively. Furthermore, *MdJMJ19* was expressed in all parts of the flower, with significantly higher expression levels in petals and sepals compared to other parts. Finally, it is noteworthy that *MdJMJ2* showed either no expression or relatively low expression levels in fruit development, shoot apex, bud, different parts of flowers, and young leaves.

3.6 Expression profiles of apple *MdPRMTs* and *MdJMJs* under different biotic and abiotic stresses

In East Asia, *Alternaria alternata* apple pathotype (AAAP) is one of the main pathogens causing apple *Alternaria* blotch disease,

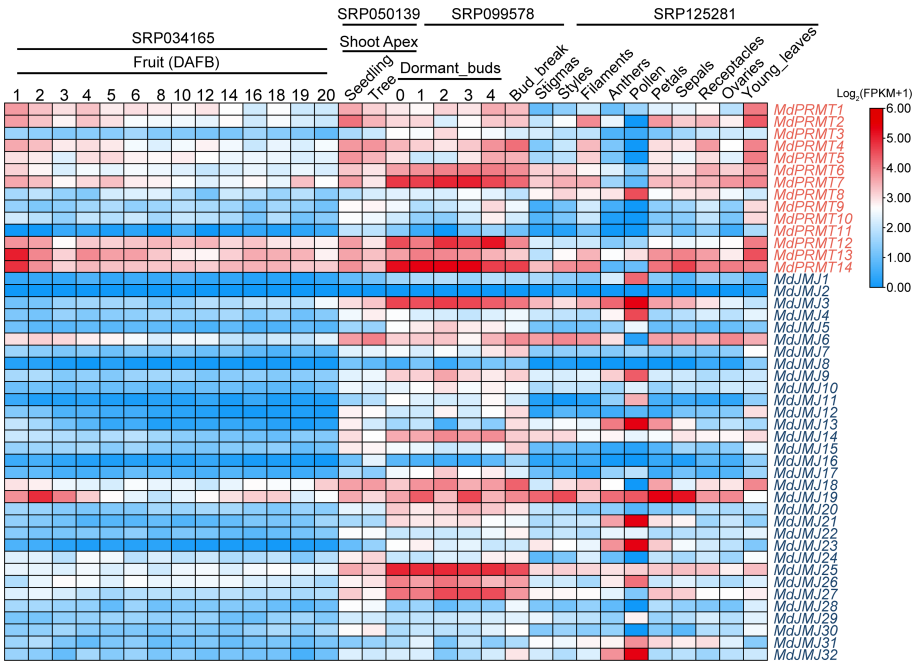


FIGURE 6
Expression profiles of the apple *MdPRMT* and *MdJMJ* genes in various organs, tissues and developmental stages. Different shades of red and blue denote the extent of the expression values according to the color bar provided [$\log_2$(FPKM+1)].

significantly affecting the growth and development of apple trees and fruit yield (Abe et al., 2010). Some *MdPRMTs* and *MdJMJs* showed decreased expression levels after infection, such as *MdJMJ6*, *MdJMJ15*, and *MdJMJ28*. Some genes showed an upregulation, such as *MdPRMT1*, *MdPRMT12* and *MdJMJ25*. In addition, certain genes demonstrated fluctuating expression levels, such as *MdPRMT14* and *MdPRMT7*, which increased in expression level from 0 to 18 hours after infection and then decreased below the expression level at 0 hours of infection. On the other hand, *MdJMJ27* and *MdJMJ3* showed a decrease in expression level at the early stage of infection and then an increase in expression level after 18 hours of infection. It is worth noting that the expression level of *MdJMJ13* was significantly higher at 18 and 72 hours after infection compared to other time points.

Pythium ultimum is one of the main pathogens causing Apple Replant Disease (ARD), which leads to growth inhibition and even death of apple seedlings when planted in orchard soil previously used for apple (or closely related species) cultivation (Zhu and Saltzgiver, 2020). The expression levels of *MdJMJ* and *MdPRMT* genes generally do not show significant changes after infection. However, it should be noted that *MdJMJ13* expression is higher than the control after 4 hours of infection, while *MdJMJ18* expression is lower than the control.

ARD causes severe growth and developmental obstacles in apple trees, and previous studies have determined *Fusarium solani* as the main pathogen causing ARD (Liu et al., 2022). There was a significant downregulation of *MdJMJ3*, *MdJMJ18*, *MdJMJ25*, *MdJMJ27*, *MdJMJ29* and *MdPRMT8*, while *MdPRMT2* showed an upregulation after *F.solani* infection. These results suggest that the above genes may play a role in apple resistance against *F. solani* infection.

Apple fruit ring rot (FRR) is a highly destructive disease caused by *Botryosphaeria dothidea*, which severely affects the apple industry's development in the Asian region (Shen et al., 2019). The results revealed that, regardless of susceptible or resistant varieties, the expression levels of *MdPRMTs* and *MdJMJs* showed no significant changes compared to the corresponding mock after *Botryosphaeria dothidea* infection (Figure 7A).

First, in terms of abiotic stress, we focused on salt stress (SRP229388) (Figure 6B). Transcriptional data from apple seedlings subjected to two types of salt stress, KCl and NaCl, showed that the expression levels of most *MdPRMTs* and *MdJMJs* were relatively low in the leaves. It is worth noting that *MdJMJ6* showed a downregulation trend in the leaves under both types of salt stress. Conversely, *MdJMJ3*, *MdJMJ13*, *MdJMJ19*, and *MdPRMT7* exhibited significant upregulation under both KCl and NaCl salt stress, with *MdJMJ3* showing higher expression levels during the initial 1 hour of KCl salt stress, followed by a significant decrease after 6 hours, while the opposite trend was observed under NaCl salt stress. In the root system, *MdJMJ13* displayed an obvious upregulation trend under both KCl and NaCl salt stress, with a significant increase in expression levels after 6 hours of NaCl stress, much higher than the expression levels under the same period of KCl stress. *MdPRMT12* showed a downregulation trend, and *MdJMJ19* exhibited an initial upregulation followed by a decrease in expression levels. A comparison between the leaf and root systems also revealed that the expression levels of *MdJMJ13*, *MdPRMT2*, *MdPRMT13*, and *MdPRMT14* under both KCl and NaCl salt stress were significantly higher in the root system than in the leaf system, whereas the opposite was observed in *MdJMJ6* and *MdPRMT7*. Neither *MdJMJ16* nor *MdJMJ2* showed expression.

In the northern region of China, low-temperature freeze injury is one of the common diseases in apple during winter and spring (Yang et al., 2011). We analyzed the RNA-seq transcriptomes of barks (epidermis, phloem, and cambium) from one-year-old

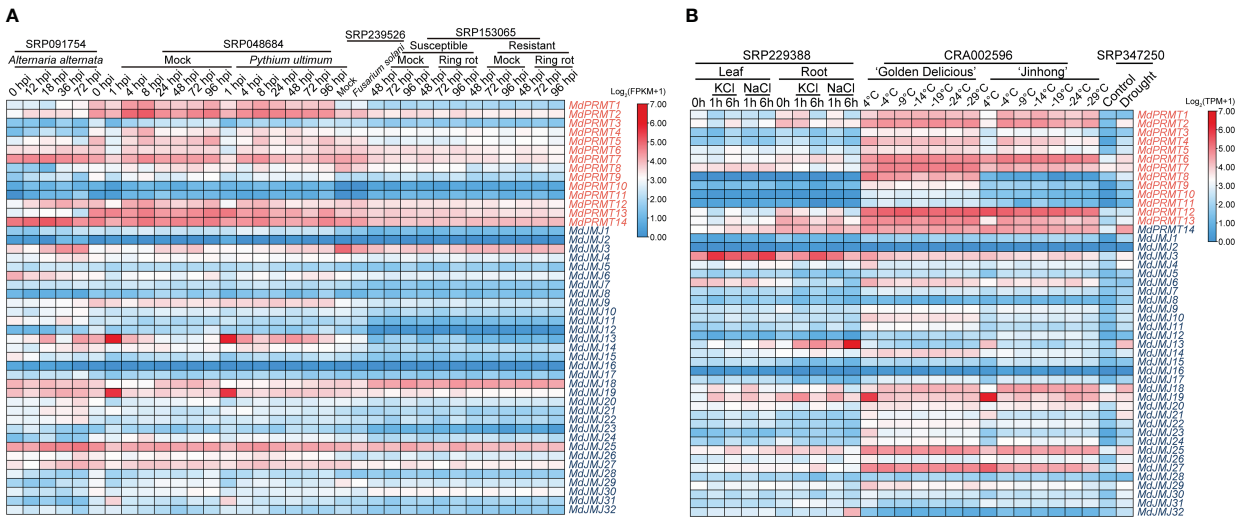


FIGURE 7
Expression profiles of the apple *MdPRMT* and *MdJMJ* genes under stress treatment. **(A)** Expression profiles of *MdPRMT* and *MdJMJ* genes at different time points under biotic stress (*Alternaria alternata*, *Pythium ultimum*, *Fusarium solani*, Ring rot) (SRP091754, SRP048684, SRP239526, SRP153065). **(B)** Expression profiles of *MdPRMT* and *MdJMJ* genes in leaves and roots at different time points under KCl and NaCl salt treatment (SRP229388), RNA-seq transcriptomes of barks (epidermis, phloem, and cambium) from one-year-old branches of two apple cultivars ('Golden Delicious' and 'Jinhong') under chilling and freezing treatments (CRA002596), and 'GL3' under drought stress (SRP347250). Different shades of red and blue denote the extent of the expression values according to the color bar provided [$\log_2$(FPKM+1) and $\log_2$(TPM+1)].

branches of two apple cultivars, 'Golden Delicious' (non-cold-tolerant varieties) and 'Jinhong' (cold-tolerant varieties) under chilling and freezing treatments (CRA002596) (Figure 7B). The results showed that the expression level of *MdPRMT8* was significantly higher in 'Golden Delicious' than in 'Jinhong,' and with the continuous decrease in temperature, its expression level showed a significant downregulation. *MdPRMT9*, *MdPRMT10*, *MdPRMT11*, and *MdJMJ23* had higher average expression levels in 'Golden Delicious' than in 'Jinhong', but their expression levels did not show an obvious pattern of change. It is worth noting that *MdJMJ19* had very high expression levels at 4°C in both apple cultivars, and as the temperature decreased, its expression levels were significantly downregulated. In particular, *MdJMJ19* in 'Golden Delicious' showed a stable expression level at -9°C and even lower temperatures, while in 'Jinhong', a slight upregulation was observed starting from -4°C, and although the expression levels decreased at -14°C and -29°C, the overall trend was still upregulation. In 'Jinhong', the expression level of *MdJMJ15* was significantly higher at -4°C and lower temperatures than at 4°C, while in 'Golden Delicious', the change in expression levels was not significant. *MdPRMT2* and *MdPRMT13* had significantly higher expression levels at -4°C in 'Jinhong' than at 4°C, and their expression levels were downregulated at lower temperatures, which was not observed in 'Golden Delicious'. *MdJMJ18* showed an upregulation of expression levels between -4°C and -14°C in both cultivars, and its expression levels decreased at even lower temperatures, indicating that this gene may play a role in apple's resistance to low temperature.

Drought is also one of the main types of stress faced during apple cultivation, especially in drought-prone areas of China (Berger et al., 2016). We analyzed the RNA-seq data of 'GL3' under drought treatment (SRP347250) (Figure 7B). The analysis results showed that compared to the control group, most *MdPRMTs* and *MdJMJs* showed varying degrees of upregulation under drought stress, with *MdPRMT2*, *MdPRMT6*, *MdPRMT14*, *MdJMJ18*, *MdJMJ13*, *MdJMJ20*, *MdJMJ21*, *MdJMJ25* and *MdJMJ27* exhibiting significant upregulation. Therefore, it can be inferred that these genes may be closely related to apple's resistance to drought.

We selected five *MdJMJ* and *MdPRMT* genes that were significantly upregulated under drought stress for qRT-PCR analysis. The results showed a significant and varying upregulation in all five genes upon PEG (as drought) treatment (Figure 8A). When two *MdJMJ* genes (*MdJMJ13* and *MdJMJ32*) were subjected to salt stress, both exhibited a decrease in expression after 6 h, followed by an increase at 12 h (Figure 8B). Similarly, the relative expression levels of the four genes under 4°C cold stress displayed a similar pattern, with an initial increase at 6 h and subsequent decline (Figure 8C).

3.7 Expression profiles of the apple *MdPRMTs* and *MdJMJs* in response to hormone treatment

Plant hormones play a crucial role in the growth and development of plants. 1-MCP (1-Methylcyclopropene) is commonly used for preserving ethylene-producing or ethylene-sensitive fruits, effectively delaying fruit senescence (Berger et al., 2016). We analyzed the RNA-seq data (SRP334206) of three apple fruit varieties ('Golden Delicious', 'Granny Smith', and 'Fuji') treated with 1-MCP (Figure 9). The analysis revealed that *MdPRMT2*, *MdPRMT7*, *MdPRMT12*, *MdPRMT13*, and *MdPRMT14* showed significantly upregulated expression in the 1-MCP treatment group compared to the control group in all three apple varieties. Additionally, the expression of *MdPRMT8* in 'Fuji' was significantly lower than the other two varieties. Similarly, the expression of *MdJMJ3*, *MdJMJ18*, *MdJMJ25*, and *MdJMJ27* was significantly higher in the 1-MCP treatment group, with *MdJMJ3* showing the most noticeable upregulation at 60 d after 1-MCP treatment. This suggests that the aforementioned genes may be important for regulating the response of apple fruits to 1-MCP.

GA_3 and NAA are important plant hormones involved in various plant growth and development processes (Hennerty and Forshey, 1972; Pharis and King, 1985). We analyzed the RNA-seq data (SRP185711) of the main floral organs of 'Honeycrisp' apples after 18 and 132 days of NAA and GA_3 treatment. The analysis showed that the expression of *MdJMJ6* in the hypanthium, ovary, and ovule was significantly lower after 18 days of NAA and GA_3 treatment compared to the hand-pollinated control (HP). Additionally, after 132 days of GA_3 treatment, all three organs showed lower expression levels of *MdJMJ6*. *MdPRMT8* and *MdJMJ19* exhibited significant upregulation in expression after 18 days of NAA and GA_3 treatment compared to HP. Furthermore, the expression levels of *MdPRMTs* and *MdJMJs* in all organs were significantly lower after 132 days of GA_3 treatment compared to 18 days (Figure 10).

Finally, we also investigated the response of apple plants at different growth and development stages to Indole-3-butyric acid (IBA), which promotes adventitious root formation (Supplementary Figure S1). The ability of apple plants to generate adventitious roots varies at different stages of growth and development. We analyzed the RNA-seq data (SRP330812) of adult and juvenile apple plants treated with IBA. Most of the *MdPRMTs* and *MdJMJs* showed no significant sensitivity to IBA treatment. However, it should be noted that in adult plants, the expression of *MdJMJ19* was upregulated after 6 and 12 hours of IBA treatment compared to mock, followed by a decrease with longer treatment time. In juvenile plants, except for the 12-hour treatment, the expression of *MdJMJ19* was higher than mock.

4 Discussion

Epigenetics involves stable heritable variations in organisms without changes in DNA sequences, including histone modifications (Berger et al., 2009). Histone modifications play roles in maintaining genome stability, gene regulation, and cellular processes. Research on histone modifications in important crops like rice (Ahmad et al., 2011) and tomato (Li et al., 2020) has shown their impact on gene expression and cellular activities. In this study, *MdJMJ* and *MdPRMT* families in apples were identified, characterized, and analyzed for potential roles in stress response, hormone regulation, and tissue development (Figure 11).

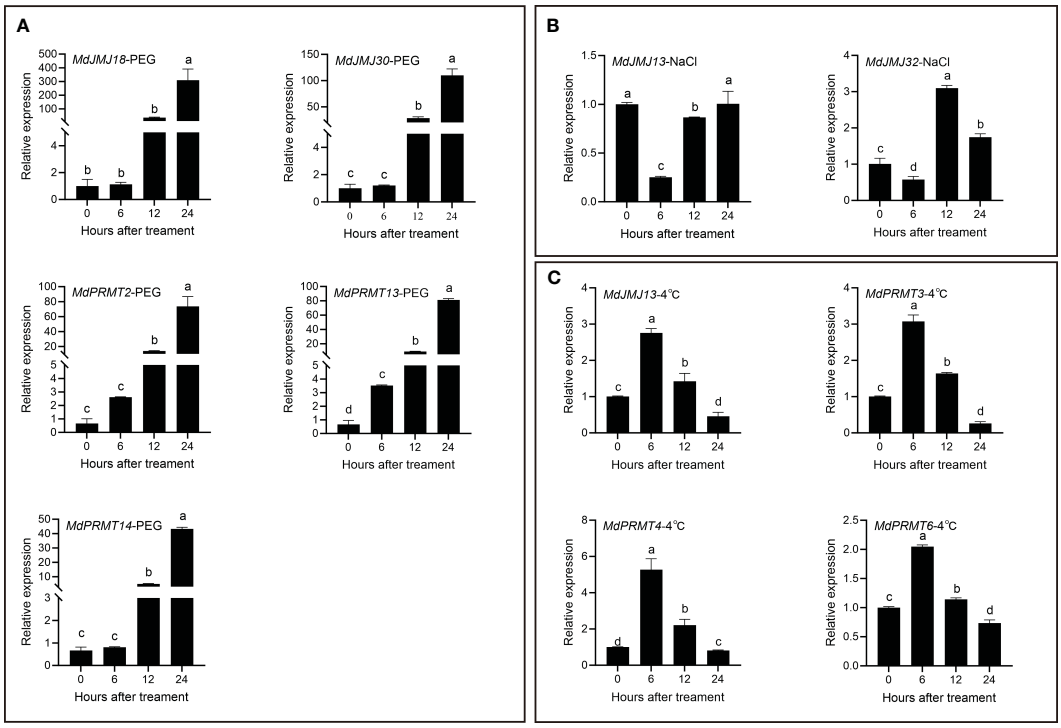


FIGURE 8
qRT-PCR analysis of selected *MdJMJ* and *MdPRMT* genes expression under PEG **(A)**, NaCl **(B)** and 4°C **(C)** treatment. 25 d-old leaves were treated with PEG, NaCl and 4°C for 0 h, 6 h, 12 h and 24 h The mean values ± SEM are shown for three biological replicates. Different letters above the bars indicate significant differences according to Duncan-test ($P < 0.05$).

In order to better understand the phylogenetic relationship between the *JMJ* and *PRMT* gene families in apple and *Arabidopsis*, we constructed a phylogenetic tree. Based on the results of related studies in *Arabidopsis*, we classified the *JMJ* gene family into corresponding subfamilies, and found that the *JMJ* genes clustered in a logical manner. Similarly, the *PRMT* genes also showed reasonable clustering. Gene duplication plays a crucial role in species evolution (Panchy et al., 2016). In the current apple

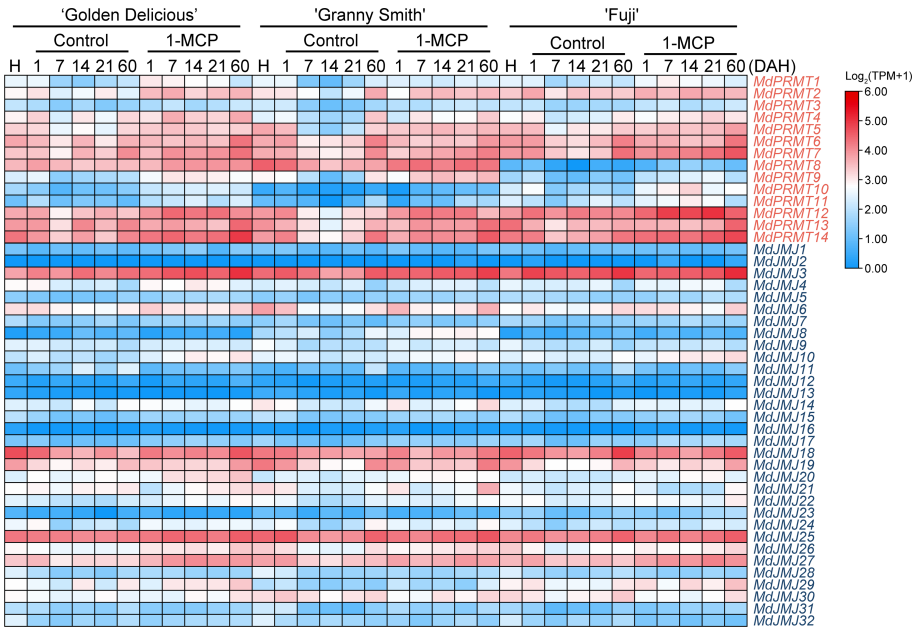


FIGURE 9
Expression profiles of the apple *MdPRMT* and *MdJMJ* genes during fruit response to 1-MCP treatment (SRP334206). Different shades of red and blue denote the extent of the expression values according to the color bar provided [log_2(TPM+1)].

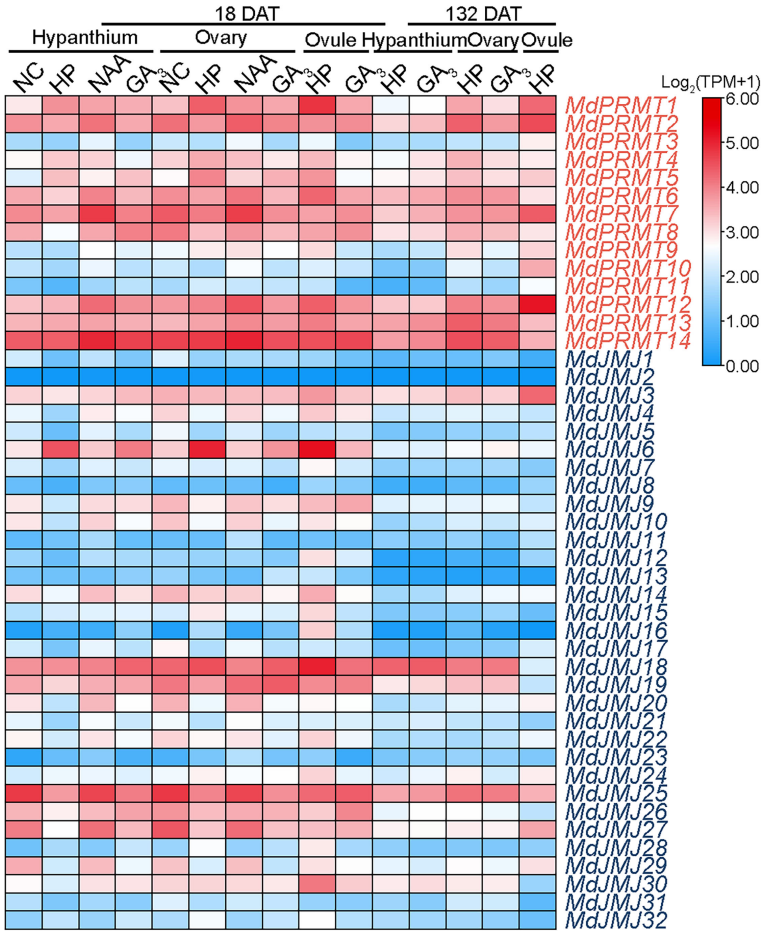


FIGURE 10
Expression profiles of the *MdPRMT* and *MdJMJ* genes in response to flower treatments with NAA and GA_3 in apple (SRP185711). Different shades of red and blue denote the extent of the expression values according to the color bar provided [log_2(TPM+1)].

reference genome, we observed that most of the *MdJMJ* and *MdPRMT* genes have undergone gene duplication. In this study, we used the *Arabidopsis* genome as an outgroup and employed Dupgen_finder to identify the types of gene duplication within the *MdJMJ* and *MdPRMT* gene families. The results indicated that apple has recently experienced a whole-genome duplication event (Qiao et al., 2019), the predominant types of gene duplications in the *MdJMJ* and *MdPRMT* families following whole-genome duplication were WGD, with a smaller proportion of PD, TRD and TD types. Generally, different categories of gene duplication exhibit distinct patterns of functional evolution. The Ka/Ks values for most homologous gene pairs in the *MdJMJ* and *MdPRMT* gene families were less than 1, suggesting they have undergone purifying selection, eliminating deleterious mutations. Notably, a cluster of *MdPRMT* genes was identified on apple chromosome 15. Analysis showed that these genes were mainly involved in proximal repeats. In the promoter regions, significant differences in the types of *cis*-regulatory elements were observed between the PD-pair (*MdPRMT9-MdPRMT11*, *MdPRMT9-MdPRMT10*). In terms of expression profiles, *MdPRMT9* exhibited higher expression levels in most parts of apple flowers compared to *MdPRMT10* and *MdPRMT11*, under various treatments including biotic stress, cold stress and 1-MCP treatment. This suggests that TD-pairs retain higher conservation in their structural homogeneity, without affecting their functional differentiation (Qiao et al., 2019), while the WGD duplication serves as the main driver for the expansion of the *MdJMJ* and *MdPRMT* gene families, contributing to the diversity in family member structures (Feng et al., 2024).

We particularly focused on the differences in *cis*-regulatory elements between TD and PD repeats in the *MdJMJ* and *MdPRMT* gene families. For example, between the TD-pair (*MdJMJ1-MdJMJ2*), 14 types of cis-acting elements were identified, but only 7 of these were shared by both genes. A similar situation was also observed between the two PD-pairs in the *MdPRMT* family. This suggests that tandem duplication and proximal duplication play a significant role in the evolution of novel functions in the *MdJMJ* and *MdPRMT* gene families (Cannon et al., 2004). Furthermore, a complementary pattern of *cis*-regulatory elements is observed in the promoter regions of WGD gene pairs. For example, *MdJMJ20* and *MdJMJ26* form a WGD duplicate pair with *MdJMJ9*, and among the predicted 14 *cis*-elements, only four (Light, MYB_drought, ABRE, and ARE) are shared by all three genes. The above results can be explained by the duplication-degeneration-complementation (DDC) model. The DDC model

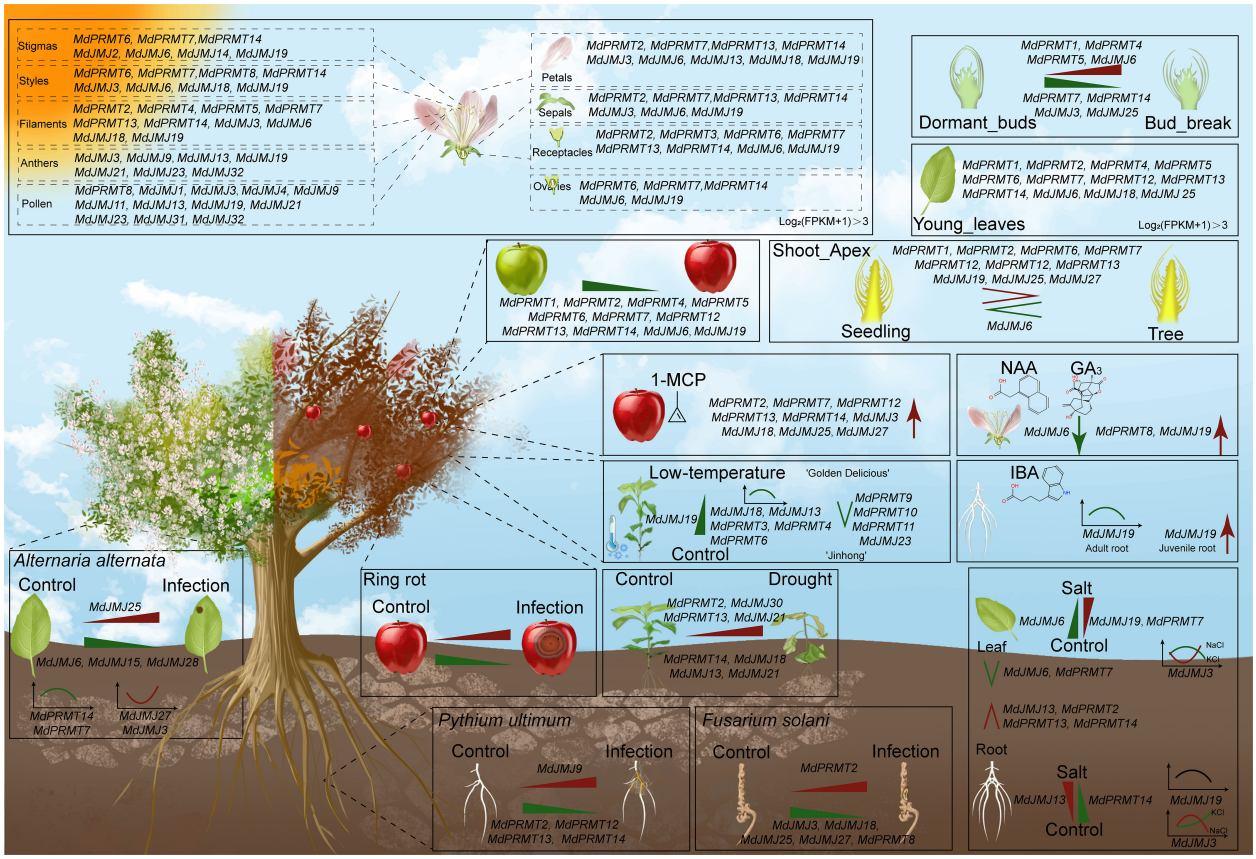


FIGURE 11
The role of the *MdJMJ* and *MdPRMT* genes in apple.

suggests that following a gene duplication event, two alleles at two different loci experience degeneration mutations at different times, leading to a pair of duplicate genes sharing the ancestral gene's functions and completing the ancestral gene's functions in a complementary manner (Hahn, 2009). Comparative analysis of gene expression in various apple tissues revealed significant tissue-specific expression differences among most duplicate gene pairs (Figure 6), suggesting that these genes have evolved by differentiating ancestral gene functions or expression patterns under different tissue conditions. Additionally, we focused on comparing *cis*-regulatory elements in the promoters of apples (*MdJMJ* and *MdPRMT* families) with those of *Arabidopsis* (*AtJMJ* and *AtPRMT* families) (Supplementary Figure S2). The average number of *cis*-elements in apples is 9.28 for *MdJMJ* and 9.57 for *MdPRMT*, compared to 7.76 for *AtJMJ* and 8.22 for *AtPRMT* in *Arabidopsis*. The differences in element types can be attributed to both the high species specificity of *cis*-elements and the result of different selection pressures and environmental conditions acting on the ancestral genes after duplication events, leading to the acquisition of new *cis*-regulatory elements in their promoters (Zhang, 2003; Lan and Pritchard, 2016). The genes in model organisms generally have more comprehensive functional studies. For example, the homologous gene of *MdJMJ21*, *AtJMJ25*, has been confirmed to be involved in the formation of the very early embryo and endosperm in *Arabidopsis* (Day et al., 2008). We found that there is an endosperm-related *cis*-acting element in the promoter region of both *MdJMJ21* and *AtJMJ25* (Supplementary Figure S2), suggesting that *MdJMJ21* may be involved in the regulation of apple endosperm formation. The *Arabidopsis* gene *AtLDL1* has been confirmed to participate in physiological (defense) and developmental (flowering time) processes (Dutta et al., 2017), possessing ARE, ABRE, and Stress_defense-related regulatory elements. Homologous to *AtLDL1*, *MdJMJ3* has eight ABRE and one ARE regulatory elements and shows up-regulated expression under drought stress treatment, and it is significantly expressed in anthers and pollen. Therefore, *MdJMJ3* may similarly participate in the regulation of apple abiotic stress and flower development processes. Further homologous analysis of some genes in the *MdJMJ* and *MdPRMT* gene families with model organisms can reveal their potential functions.

Multiple models have been proposed to elucidate the retention and evolution of gene duplications, including gene dosage balance (Birchler et al., 2005), subfunctionalization (SF), neofunctionalization (NF) (Sandve et al., 2018), expression specialization, and pseudogenization (Qiao et al., 2019). Based on expression profiles, we studied the expression differences between duplicated genes and explored the mechanisms of gene retention. Interestingly, at different stages of fruit development, many gene pairs showed complementary expression patterns, which may indicate subfunctionalization. For instance, the gene pair *MdPRMT1* and *MdPRMT12*, where one gene

copy had higher expression levels at several stages while the other gene copy had higher expression levels at the remaining stages. Some gene pairs exhibited parallel expression patterns, suggesting that gene dosage balance is imposed in the evolution of duplicated gene pairs to maintain the total expression level of ancestral genes. For example, duplicated gene pairs like *MdPRMT9-MdPRMT10* and *MdJMJ1-MdJMJ2* showed similar expression patterns at different stages of fruit development. Additionally, some gene pairs showed expression specialization and nonfunctionalization, such as *MdJMJ19-MdJMJ31* and *MdJMJ18-MdJMJ30*, where one gene copy was highly expressed at almost all stages of dormant buds, while the other gene copy showed low or no expression. We primarily focused on the performance of *MdJMJ* and *MdPRMT* under abiotic stress. It was found that some duplication gene pairs exhibited similar expression patterns under abiotic stress treatments, such as *MdJMJ18-MdJMJ30* and *MdJMJ13-MdJMJ32*, where qPT-PCR and transcriptome analyses confirmed their consistent expression patterns under drought stress and salt stress, and a certain number of related *cis*-regulatory elements were identified in their promoter regions, possibly involved in regulating gene expression under stress conditions. These results demonstrate the complexity of the expansion and evolution of the *MdJMJ* and *MdPRMT* gene families, laying the foundation for further elucidating their more specific regulatory mechanisms.

5 Conclusions

In this study, a total of 14 *MdPRMT* genes and 32 *MdJMJ* genes were identified in apples. Co-linearity analysis showed that both gene families exhibited high conservation between *Arabidopsis* and apple. Promoter analysis indicated that *MdPRMTs* and *MdJMJs* may play important roles in plant growth and development, light response, hormone response, and stress response. Based on the analysis of transcription levels of *MdPRMTs* and *MdJMJs* in different tissues and developmental stages, we found that *MdPRMTs* and *MdJMJs* may have multiple functions in the growth and development process of apples. Especially, *MdPRMT13*, *MdPRMT14*, and *MdJMJ19* showed a downtrend during fruit ripening and aging process. *MdPRMT7*, *MdPRMT12*, *MdJMJ3*, *MdJMJ13*, *MdJMJ21*, *MdJMJ23*, *MdJMJ25*, *MdJMJ26*, *MdJMJ27*, and *MdJMJ32* exhibited high expression levels in specific organs and tissues of apples. *MdJMJ25*, *MdPRMT7*, *MdPRMT14*, *MdJMJ9*, *MdPRMT2*, *MdJMJ3*, *MdJMJ18*, *MdJMJ27*, and *MdPRMT8* were involved in response to biotic stress. *MdPRMT2*, *MdPRMT7*, *MdPRMT12*, *MdPRMT13*, and *MdPRMT14* showed a significant upregulation in the 1-MCP-treated group of three apple varieties. *MdPRMT8* and *MdJMJ19* exhibited significant upregulation within 18 days after NAA and GA_3 treatment. *MdJMJ19* may be associated with apple response to IBA.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/, SRP034165, SRP050139, SRP0999578, SRP125281, SRP334206, SRP185711, SRP229388, SRP347250, SRP091754, SRP048684, SRP239526, SRP153065, SRP330812, https://ngdc.cncb.ac.cn/, CRA002596.

Author contributions

SS: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. MJ: Data curation, Formal analysis, Investigation, Software, Validation, Writing – review & editing. JC: Formal analysis, Investigation, Validation, Writing – review & editing. MZ: Investigation, Software, Validation, Writing – review & editing. XX: Conceptualization, Funding acquisition, Resources, Writing – review & editing. CC: Conceptualization, Formal analysis, Project administration, Resources, Software, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was funded by National Natural Science Foundation of China (grant numbers 32172523); The fruit innovation team of Shandong modern agricultural industry technology system (SDAIT-06-05), and the Shandong Key R&D Plan (Agricultural Variety Project) (No. 2022LZGCQY008).

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2024.1381753/full#supplementary-material

SUPPLEMENTARY TABLE 4
Orthologous gene pairs between apple, Arabidopsis, rice and zea mays JMJ and PRMT Genes.

References

Abe, K., Iwanami, H., Kotoda, N., Moriya, S., and Takahashi, S. (2010). Evaluation of apple genotypes and *Malus* species for resistance to Alternaria blotch caused by *Alternaria alternata* apple pathotype using detached-leaf method. *Plant Breed.* 129, 208–218. doi: 10.1111/j.1439-0523.2009.01672.x

Ahmad, A., Dong, Y., and Cao, X. (2011). Characterization of the *PRMT* gene family in rice reveals conservation of arginine methylation. *PloS One* 6, e22664. doi: 10.1371/journal.pone.0022664

Alvarez-Venegas, R. (2005). Methylation patterns of histone H3 Lys 4, Lys 9 and Lys 27 in transcriptionally active and inactive *Arabidopsis* genes and in *atx1* mutants. *Nucleic Acids Res.* 33, 5199–5207. doi: 10.1093/nar/gki830

Anand, R., and Marmorstein, R. (2007). Structure and mechanism of lysine-specific demethylase enzymes. *J. Biol. Chem.* 282, 35425–35429. doi: 10.1074/jbc.R700027200

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

Bedford, M. T., and Richard, S. (2005). Arginine methylation:an emerging regulator of protein function. *Mol. Cell* 18, 263–272. doi: 10.1016/j.molcel.2005.04.003

Berger, S. L., Kouzarides, T., Shiekhattar, R., and Shilatifard, A. (2009). An operational definition of epigenetics. *Genes Dev.* 23, 781–783. doi: 10.1101/gad.1787609

Berger, J., Palta, J., and Vadez, V. (2016). Review: An integrated framework for crop adaptation to dry environments: responses to transient and terminal drought. *Plant Sci.* 253, 58–67. doi: 10.1016/j.plantsci.2016.09.007

Birchler, J. A., Riddle, N. C., Auger, D. L., and Veitia, R. A. (2005). Dosage balance in gene regulation: biological implications. *Trends Genet.* 21, 219–226. doi: 10.1016/j.tig.2005.02.010

Bray, N. L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* 34, 525–527. doi: 10.1038/nbt.3519

Cannon, S. B., Mitra, A., Baumgarten, A., Young, N. D., and May, G. (2004). The roles of segmental and tandem gene duplication in the evolution of large gene families in *Arabidopsis thaliana*. *BMC Plant Biol.* 4, 10. doi: 10.1186/1471-2229-4-10

Chen, C., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y., et al. (2020). TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant* 13, 1194–1202. doi: 10.1016/j.molp.2020.06.009

Copenhaver, G. P., Chen, Q., Chen, X., Wang, Q., Zhang, F., Lou, Z., et al. (2013). Structural basis of a histone H3 lysine 4 demethylase required for stem elongation in rice. *PloS Genet.* 9, e1003239. doi: 10.1371/journal.pgen.1003239

Cui, X., Jin, P., Cui, X., Gu, L., Lu, Z., Xue, Y., et al. (2013). Control of transposon activity by a histone H3K4 demethylase in rice. *Proc. Natl. Acad. Sci.* 110, 1953–1958. doi: 10.1073/pnas.1217020110

Da, L., Liu, Y., Yang, J., Tian, T., She, J., Ma, X., et al. (2019). AppleMDO: A multi-dimensional omics database for apple co-expression networks and chromatin states. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.01333

Day, R. C., Herridge, R. P., Ambrose, B. A., and Macknight, R. C. (2008). Transcriptome analysis of proliferating *Arabidopsis* endosperm reveals biological implications for the control of syncytial division, cytokinin signaling, and gene expression regulation. *Plant Physiol.* 148, 1964–1984. doi: 10.1104/pp.108.128108

Dutta, A., Choudhary, P., Caruana, J., and Raina, R. (2017). *JMJ27*, an *Arabidopsis* H3K9 histone demethylase, modulates defense against pseudomonas syringe and flowering time. *Plant J.* 91, 1015–1028. doi: 10.1111/tpj.13623

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39, 783–791. doi: 10.2307/2408678

Feng, X., Chen, Q., Wu, W., Wang, J., Li, G., Xu, S., et al. (2024). Genomic evidence for rediploidization and adaptive evolution following the whole-genome triplication. *Nat. Commun.* 15, 1635. doi: 10.1038/s41467-024-46080-7

Genschik, P., Marrocco, K., Bach, L., Noir, S., and Criqui, M.-C. (2014). Selective protein degradation: a rheostat to modulate cell-cycle phase transitions. *J. Exp. Bot.* 65, 2603–2615. doi: 10.1093/jxb/ert426

Hahn, M. W. (2009). Distinguishing among evolutionary models for the maintenance of gene duplicates. *J. Hered.* 100, 605–617. doi: 10.1093/jhered/esp047

Hennerty, M. J., and Forshey, C. G. (1972). Tree physiological condition as a source of variation in chemical thinning of apple fruits. *HortScience* 7, 259–260. doi: 10.21273/HORTSCI.7.3.259

Hurst, L. D. (2002). The Ka/Ks ratio: diagnosing the form of sequence evolution. *Trends Genet.* 18, 486. doi: 10.1016/S0168-9525(02)02722-1

Kenchanmane Raju, S. K., Ledford, M., and Niederhuth, C. E. (2023). DNA methylation signatures of duplicate gene evolution in angiosperms. *Plant Physiol.* 192, 2883–2901. doi: 10.1093/plphys/kiad220

Kouzarides, T. (2007). Chromatin modifications and their function. *Cell* 128, 693–705. doi: 10.1016/j.cell.2007.02.005

Lan, X., and Pritchard, J. K. (2016). Coregulation of tandem duplicate genes slows evolution of subfunctionalization in mammals. *Science* 352, 1009–1013. doi: 10.1126/science.aad8411

Lawrence, M., Daujat, S., and Schneider, R. (2016). Lateral thinking: how histone modifications regulate gene expression. *Trends Genet.* 32, 42–56. doi: 10.1016/j.tig.2015.10.007

Lescot, M., Déhais, P., Thijs, G., Marchal, K., Moreau, Y., Van de Peer, Y., et al. (2002). PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res.* 30, 325–327. doi: 10.1093/nar/30.1.325

Li, Z., Jiang, G., Liu, X., Ding, X., Zhang, D., Wang, X., et al. (2020). Histone demethylase *SlJMJ6* promotes fruit ripening by removing H3K27 methylation of ripening-related genes in tomato. *New Phytol.* 227, 1138–1156. doi: 10.1111/nph.16590

Li, W., Yan, J., Wang, S., Wang, Q., Wang, C., Li, Z., et al. (2021). Genome-wide analysis of SET-domain group histone methyl transferases in apple reveals their role in development and stress responses. *BMC Genomics* 22, 283. doi: 10.1186/s12864–021-07596–0

Liu, Y., Liu, Q., Li, X., Tang, Z., Zhang, Z., Gao, H., et al. (2022). Exogenous dopamine and *MdTyDC* over expression enhance apple resistance to fusarium solani. *Phytopathology* 112, 2503–2513. doi: 10.1094/PHYTO-04-22-0142-R

Liu, C., Lu, F., Cui, X., and Cao, X. (2010). Histone methylation in higher plants. *Annu. Rev. Plant Biol.* 61, 395–420. doi: 10.1146/annurev.arplant.043008.091939

Luo, M., Hung, F.-Y., Yang, S., Liu, X., and Wu, K. (2014). Histone lysine demethylases and their functions in plants. *Plant Mol. Biol. Rep.* 32, 558–565. doi: 10.1007/s11105-013-0673-1

Ma, N., Tan, H., Liu, X., Xue, J., Li, Y., and Gao, J. (2006). Transcriptional regulation of ethylene receptor and *CTR* genes involved in ethylene-induced flower opening in cut rose (*Rosa hybrida*) cv. Samantha. *J. Exp. Bot.* 57, 2763–2773. doi: 10.1093/jxb/erl033

Niu, L., Lu, F., Pei, Y., Liu, C., and Cao, X. (2007). Regulation of flowering time by the protein arginine methyl transferase *AtPRMT10*. *EMBO Rep.* 8, 1190–1195. doi: 10.1038/sj.embor.7401111

Niu, L., Zhang, Y., Pei, Y., Liu, C., and Cao, X. (2008). Redundant requirement for a pair of PROTEIN ARGININE METHYLTRANSFERASE4 homologs for the proper regulation of *Arabidopsis* flowering time. *Plant Physiol.* 148, 490–503. doi: 10.1104/pp.108.124727

Panchy, N., Lehti-Shiu, M., and Shiu, S. H. (2016). Evolution of gene duplication in plants. *Plant Physiol.* 171, 2294–2316. doi: 10.1104/pp.16.00523

Pei, Y., Niu, L., Lu, F., Liu, C., Zhai, J., Kong, X., et al. (2007). Mutations in the type II protein arginine methyltransferase *AtPRMT5* result in pleiotropic developmental defects in *Arabidopsis*. *Plant Physiol.* 144, 1913–1923. doi: 10.1104/pp.107.099531

Perini, P., Pasquali, G., Margis-Pinheiro, M., de Oliviera, P. R. D., and Revers, L. F. (2014). Reference genes for transcriptional analysis of flowering and fruit ripening stages in apple (*Malus* × *domestica* Borkh.). *Mol. Breed.* 34, 829–842. doi: 10.1007/s11032-014-0078-3

Pharis, R. P., and King, R. W. (1985). Gibberellins and reproductive development in seed plants. *Annu. Rev. Plant Biol.* 36, 517–568. doi: 10.1146/annurev.pp.36.060185.002505

Qiao, X., Li, Q., Yin, H., Qi, K., Li, L., Wang, R., et al. (2019). Gene duplication and evolution in recurring polyploidization-diploidization cycles in plants. *Genome Biol.* 20, 38. doi: 10.1186/s13059-019-1650-2

Rohde, A., and Bhalerao, R. P. (2007). Plant dormancy in the perennial context. *Trends Plant Sci.* 12, 217–223. doi: 10.1016/j.tplants.2007.03.012

Sandve, S. R., Rohlfs, R. V., and Hvidsten, T. R. (2018). Subfunctionalization versus neofunctionalization after whole-genome duplication. *Nat. Genet.* 50, 908–909. doi: 10.1038/s41588-018-0162-4

Shen, F., Huang, Z., Zhang, B., Wang, Y., Zhang, X., Wu, T., et al. (2019). Mapping gene markers for apple fruit ring rot disease resistance using a multi-omics approach. *G3-Genes Genomes Genet.* 9, 1663–1678. doi: 10.1534/g3.119.400167

Sun, Q., and Zhou, D.-X. (2008). Rice jmjC domain-containing gene *JMJ706* encodes H3K9 demethylase required for floral organ development. *Proc. Natl. Acad. Sci. United States America* 105, 13679–13684. doi: 10.1073/pnas.0805901105

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

Thomas Jenuwein, C. D. A. (2001). Translating the histone code. *Science* 293, 1074–1080. doi: 10.1126/science.1063127

Wang, H., Huang, Z.-Q., Xia, L., Feng, Q., Erdjument-Bromage, H., Strahl, B. D., et al. (2001). Methylation of histone H4 at arginine 3 facilitating transcriptional activation by nuclear hormone receptor. *Am. Assoc. Advance. Sci.* 293, 853–857. doi: 10.1126/science.1060781

Xu, G., Guo, C., Shan, H., and Kong, H. (2012). Divergence of duplicate genes in exon–intron structure. *Proc. Natl. Acad. Sci.* 109, 1187–1192. doi: 10.1073/pnas.1109047109

Yang, W., Liu, X. D., Chi, X. J., Wu, C. A., Li, Y. Z., Song, L. L., et al. (2011). Dwarf apple *MbDREB1* enhances plant tolerance to low temperature, drought, and salt stress via both ABA-dependent and ABA-independent pathways. *Planta* 233, 219–229. doi: 10.1007/s00425-010-1279-6

Zhang, J. (2003). Evolution by gene duplication: an update. *Trends Ecol. Evol.* 18, 292–298. doi: 10.1016/S0169-5347(03)00033-8

Zhao, J., Guo, R., Guo, C., Hou, H., Wang, X., and Gao, H. (2016). Evolutionary and expression analyses of the apple basic leucine zipper transcription factor family. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.00376

Zhu, Y., and Saltzgiver, M. (2020). A systematic analysis of apple root resistance traits to *Pythium ultimum* infection and the underpinned molecular regulations of defense activation. *Horticult. Res.* 7, 62. doi: 10.1038/s41438-020-0286-4

frontiers | Frontiers in Plant Science

TYPE Mini Review
PUBLISHED 06 June 2024
DOI 10.3389/fpls.2024.1407789

Check for updates

OPEN ACCESS

EDITED BY
Reyazul Rouf Mir,
Sher-e-Kashmir University of Agricultural Sciences and Technology, India

REVIEWED BY
Vikas Venu Kumaran,
Indian Agricultural Research Institute (ICAR), India
Usman Aziz,
Northwestern Polytechnical University, China
Mohd. Anwar Khan,
Sher-e-Kashmir University of Agricultural Sciences and Technology of Kashmir, India

*CORRESPONDENCE
Sajad Ali
sajadmicro@yu.ac.kr
Mohammed A. Almalki
malmalki@kfu.edu.sa
Zahoor Ahmad Mir
zahoorbio@gmail.com

†These authors have contributed equally to this work

RECEIVED 27 March 2024
ACCEPTED 16 May 2024
PUBLISHED 06 June 2024

CITATION
Tyagi A, Ali S, Mir RA, Sharma S, Arpita K, Almalki MA and Mir ZA (2024) Uncovering the effect of waterlogging stress on plant microbiome and disease development: current knowledge and future perspectives.
Front. Plant Sci. 15:1407789.
doi: 10.3389/fpls.2024.1407789

COPYRIGHT
© 2024 Tyagi, Ali, Mir, Sharma, Arpita, Almalki and Mir. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Uncovering the effect of waterlogging stress on plant microbiome and disease development: current knowledge and future perspectives

Anshika Tyagi[1†], Sajad Ali[1*†], Rakeeb Ahmad Mir[2†], Sandhya Sharma[3†], Kumari Arpita[3], Mohammed A. Almalki[4*] and Zahoor Ahmad Mir[5*]

[1]Department of Biotechnology, Yeungnam University, Gyeongsan, Republic of Korea, [2]Department of Biotechnology, School of Life Sciences, Central University of Kashmir, Ganderbal, Jammu and Kashmir, India, [3]ICAR-National Institute for Plant Biotechnology, New Delhi, India, [4]Department of Biological Sciences, College of Science, King Faisal University, Al-Ahsa, Saudi Arabia, [5]Department of Plant Science and Agriculture, University of Manitoba, Winnipeg, MB, Canada

Waterlogging is a constant threat to crop productivity and ecological biodiversity. Plants face multiple challenges during waterlogging stress like metabolic reprogramming, hypoxia, nutritional depletion, reduction in gaseous exchange, pH modifications, microbiome alterations and disease promotion all of which threaten plants survival. Due to global warming and climatic change, the occurrence, frequency and severity of flooding has dramatically increased posing a severe threat to food security. Thus, developing innovative crop management technologies is critical for ensuring food security under changing climatic conditions. At present, the top priority among scientists is to find nature-based solutions to tackle abiotic or biotic stressors in sustainable agriculture in order to reduce climate change hazards to the environment. In this regard, utilizing plant beneficial microbiome is one of the viable nature based remedial tool for mitigating abiotic stressors like waterlogging. Beneficial microbiota provides plants multifaceted benefits which improves their growth and stress resilience. Plants recruit unique microbial communities to shield themselves against the deleterious effects of biotic and abiotic stress. In comparison to other stressors, there has been limited studies on how waterlogging stress affects plant microbiome structure and their functional traits. Therefore, it is important to understand and explore how waterlogging alters plant microbiome structure and its implications on plant survival. Here, we discussed the effect of waterlogging stress in plants and its microbiome. We also highlighted how waterlogging stress promotes pathogen occurrence and disease development in plants. Finally, we highlight the knowledge gaps and areas for future research directions on unwiring how waterlogging affects plant microbiome and its functional traits. This will pave the way for identifying resilient microbiota that can be engineered to promote their positive interactions with plants during waterlogging stress.

KEYWORDS

waterlogging, plant microbiome, hypoxia, pathogens, signaling, disease

Introduction

In recent decades, harsh environmental conditions, such as floods, drought, and extreme temperatures, have caused a significant drop in agricultural yields across the globe (Kreibich et al., 2022; Furtak and Wolińska, 2023). According to the Food and Agriculture Organization of the United Nations (FAO), an increase in food production of around 70% by 2050 is necessary to fulfill the demand of an expanding population (FAO, 2021). However, in fulfilling this demand, there is need to develop future climate resilient smart crops in sustainable agriculture. Numerous problems arise in the forecasts of what the future holds for our global community, including how to strengthen food production in light of the escalating effects of climate change and rising population. Climate change has dramatically increased the magnitude and occurrence of environmental stressors like floods which affects crop productivity and food security (Kreibich et al., 2022). Among environmental stressors, waterlogging stress has emerged as a significant threat to agricultural output because it alters not only key plant physiological and biochemical features but also alters microbiome and soil physiochemical properties (Francioli et al., 2021; Nio and Mantilen, 2023). In nature, seasonal flooding is a regular occurrence in various ecosystems and has a favorable impact on biodiversity and production. Flooding may benefit agriculture by reloading soil nutrients in floodplains, creating new homes for wildlife, and reviving wetlands (Tonkin et al., 2018). However, unanticipated and uncontrolled floods, on the other hand, are one of the most damaging natural catastrophes, with the ability to wreak huge damage not only to agriculture and but also endangers public health (Kron, 2005). Climate models indicate that flooding events may become more frequent and severe in the near future (Jongman et al, 2014). Under flooding conditions, plants can be either completely submerged or partially submerged which can have distinct impact on their physiological, biochemical and morphological traits (Sasidharan et al., 2017). In other words, flooding is classified into two types: waterlogging, which occurs when water is on the soil surface and only plant roots are submerged and submergence, in which the whole plant can be either underwater/fully immersed or partially submerged (Jia et al., 2022). In the field, waterlogging can occur quickly after a heavy rainstorm or as a result of flood, which leads plants to hypoxic conditions (Voesenek et al., 2016).

Plants under waterlogging stress becomes more susceptible to microbial pathogens which further endangers their survival (Moslemi et al., 2018). On the other hand, waterlogging also leads dramatic alteration in root microbiome which has huge impact on plants survival under unfavorable conditions (Francioli et al., 2022; Leelastwattanagul et al., 2023). In nature, plants are associated with diverse and taxonomically structured microbial communities including bacteria, fungi, and viruses, which are called the plant microbiota (Trivedi et al, 2020). There are numerous reports which highlight the importance of beneficial microbes in improving not only plant growth but also their tolerance to different stressors (Bokhari et al, 2019; Shekhawat et al, 2021; Timmusk et al., 2023). Some of the key function's microbes can assist plants are nutrient availability, modulation of growth and defense phytohormonal signaling cascades, enhances stress tolerance and soil fertility (Ali et al., 2022a, Ali et al, 2022b, Ali et al., 2023). Despite significant progress in crop cultivar genetic modification and cultivation practices that reduce waterlogging effects, the impact of rhizosphere microorganisms in plant resistance to waterlogging has received little attention. To reduce flooding stress, two primary tactics may be implemented: traditional water management facilities (e.g., drainage, dikes) and natural based solutions (Zölch et al., 2017). Also, agronomic solutions for dealing with submergence or waterlogging include creating standard models for predicting and assessing crop loss due to floods for risk management, decision making and economic insurance. Exploring plant microbiome under waterlogging stress can provide novel nature-based strategy for improving plant tolerance to waterlogging stress. Previous studies have reported the microbial inoculation can ease waterlogging induced effects in plants. For example, inoculation of plants with Bacillus sps, producing 1-aminocyclopropane-1-carboxylic acid deaminase, lower stress-induced ethylene levels, thereby protecting plants from waterlogging stress (Ali and Kim, 2018). Similarly, Farwell et al. (2007), revealed that inoculation of *Pseudomonas putida* UW4 generates ACC deaminase, which mitigates the effects of waterlogging and metal stress. Several ACC deaminase-producing bacterial strains, including *Serratia ureilytica*, *Achromobacter xylosoxidans*, *Ochrobactrum rhizosphaerae*, and *Herbaspirillum seropedicae*, were isolated from the rhizosphere of waterlogged *Ocimum sanctum* which may protect plants from waterlogging-induced damage (Compant et al., 2019). These studies further support the notion that microbiota can be an important tool for mitigating waterlogging stress in sustainable agriculture. There are many reports which have shown that microbiome as a significant component for improving plant health and resilience to environmental stressors (Koskella et al., 2017; Lau et al., 2017; Compant et al., 2019). Unlike other stresses, there have been few studies examining the influence of waterlogging on plant microbiota. There are reports that waterlogging promotes anaerobes and disease-causing pathogens which jeopardize plants survival (Hsu and Shih, 2013; Leelastwattanagul et al., 2023). However, it is likely that plants may also recruit stress-relieving microbiome, with the ability to promote or adapt to waterlogging stress necessitates future investigations. The effects of waterlogging stress on microbial diversity and plant microbiome interactions in not fully explored despite the availability of high throughput tools. This mini review offers an update on how waterlogging stress affects plants and their microbiota. First, we discuss the effect of waterlogging in plants and signaling evolved. Next, we discussed the effect of waterlogging stress on plant microbiota. We also highlight how waterlogging promotes pathogen distribution, distribution and disease severity in different plant systems.

Waterlogging stress in plants

Waterlogging affects multifaceted morphological, physiological and biochemical traits in plants which are crucial for their growth and survival. The main challenges plants face during waterlogging

stress are reduction in the rate of gas exchange, hypoxia, low nutrient absorption, preventing aerobic respiration, increased reactive oxygen species (ROS) and ethylene levels (Ashraf, 2012; Tamang et al., 2014). Waterlogging also affects root system architecture such as growth inhibition of lateral roots which is due to the interference of ethylene with local auxin signaling (Shukla et al., 2019). Waterlogging altered root system fails to transport water and nutrients to aerial parts thereby causes reduced apoplastic water movement (Sauter, 2013). Waterlogging stress also affects chemical or hydraulic signals that cause stomatal closure, eventually contributing to reduced leaf development (Li et al., 2015). Another common response to floods is a reduction in photosynthesis. On the other hand, waterlogging triggers the accumulation of toxic compounds, carbon starvation and cytoplasmic acidification which eventually leads to plant death. On the other hand, plants vary in their capacity to survive the detrimental effects of waterlogging due to their rapid or induced modifications in plant traits, like plant height, adventitious roots aerenchyma production, changes in leaf anatomy, improved shoot elongation, starch storage hyponasty, barriers against radial oxygen loss (Voesenek et al., 2016). For example, taller plants, in particular, with aerenchyma content and larger specific leaf area may retain greater levels of gas exchange during a flood and so continue to grow (Colmer et al., 2019). Furthermore, certain plants may store enormous amounts of starch in their underground structures, alter their metabolic rates, and have the ability to develop quickly after the flood waters subside (Voesenek and Bailey-Serres, 2015).

Plants under go rapid metabolic and anatomical reprogramming during waterlogging stress in order to survive (Tyagi et al., 2023). Previous research has shown that waterlogging triggers local hypoxia-driven responses in the roots as well as systemic responses in the shoots, including changes in hormonal dynamics, metabolic reprogramming, ubiquitin-dependent protein degradation, and a variety of other molecular and metabolic responses (Hsu et al., 2011). Two hormones namely Abscisic acid (ABA) and ethylene (ET) were identified as key drivers for systemic signaling during water stress (Hsu et al., 2011; TSAI et al., 2014). One of the earliest responses to anoxia conditions in plant roots or shoots is the activation of calcium and reactive oxygen species (ROS) signaling cascades and suppression of mitochondrial respiration (Yang et al., 2023). Some of the key players like vacuolar H^+/calcium transporter CATION/PROTON EXCHANGER 1 (CAX1) and the RESPIRATORY BURST OXIDASE HOMOLOGs D and F (*RBOHD* and *RBOHF*) that drive early response during waterlogging have been identified that regulates distinct anoxia response like aerenchyma formation (Liu et al., 2017; Yang et al., 2022). Recently, Peláez-Vico et al. (2023), identified GLUTAMATE-LIKE RECEPTOR 3.3 and 3.6 (GLR3.3 and GLR3.6) calcium channels, RBOHD, and aquaporin PLASMA MEMBRANE INTRINSIC PROTEIN 2,1 (PIP2,1) proteins as potential players involved in waterlogging systemic signaling in Arabidopsis. These studies provide novel insights on how plants respond to waterlogging stress at molecular level. However, future studies are required to unravel early, localized and systemic signaling form root to shoot and also the role of cell wall sensors and calcium channels, hormones and transporters during waterlogging stress, as well as how they regulate signal perception and transduction.

Plant microbe interactions under stress conditions

Plant-microbiome interactions are complex which can be beneficial or harmful in nature. For example, beneficial microbiome provides an array of benefits to plants such as nutrient availability and uptake, nitrogen fixation, promote growth, antagonism towards pathogens, boost stress resilience and improve soil fertility (Ali et al., 2023). In contrast, harmful microbiota can be saprophytic, biotrophic, hemi biotrophic and necrotrophic inhibiting or killing the host through a variety of mechanisms (Wille et al., 2019). There have been numerous studies highlighting the plant-microbe interaction mechanisms, such as how plants respond to microbial colonization and how microbial diseases and symbionts modify plant cellular processes (Cheng et al., 2021). Interestingly, plant microbiome is an essential determinant of plant health and also one of the important drivers for plant survival under stressful conditions (Compant et al., 2019). Plant beneficial microbiomes or their metabolites are often used bioinoculants or biostimulants to enhance sustainable plant development, and have emerged as a viable alternative to agrochemicals that have negative environmental and health consequences.

Plant microbiome interactions occurs at different plant compartments with distinct habitats like phyllosphere, endosphere and rhizosphere. Plant microbiome assembly is an intricate process which is highly influenced by numerous genetic, biochemical, and physiological and environmental factors. For example, microbes have different growth conditions in terms of physiology, nutrients, pH, temperature and moisture in addition to host factors all of which can have significant impact on their assembly and host interactions in above and below ground plant organ systems (Trivedi et al., 2020). Plants ability to produce diverse chemical compounds like hormones, flavonoids mucilage, and other chemicals from roots affects microbial development, attracts particular bacteria, and can vary the rhizosphere features. However, soil microbiota, on the other hand, are sensitive to environmental changes, which has significant impact on plant survival (Zhou et al., 2023). Increased climate change and other harsh environmental stressors have not only direct effect on plant growth and yield production but also on its beneficial microbiome and their interactions. As, climate change is increasing the frequency and intensity of drought, flooding, and global temperatures is rising all of which changes the composition and activity of plant microbiomes, potentially affecting host functional attributes. Many studies have shown that environmental stressors change plant microbiome which can have distinct impact on growth and adaptive traits or can be either beneficial or detrimental to their host plants. For example, under drought stress plant recruit selective drought tolerant bacterial taxa which supports their growth under drought stress (Fitzpatrick et al., 2020). Similarly, under salinity stress plant shape unique microbiome

which alleviates their salt induced effects (Xu et al., 2020). Similarly, endophytes have been found to promote seed germination during heat and drought stress (Hubbard et al., 2019). According to Wipf et al. (2021), *Sorghum bicolor* under drought and heat stress shapes particular microbiota belongs to Actinobacteria which are known to promote growth under stress conditions. On the other hand, environmental cues can have detrimental impact on plants by promoting harmful microbiota. For example, during waterlogging anaerobes and pathogens can dominate which can have detrimental impact on plant growth and survival (Moslemi et al., 2018). Previous research has demonstrated that flooding affects the root microbiome by decreasing immunological modulator beneficial bacterial communities making plants more susceptible to disease (Soltani et al., 2010; Kavamura et al., 2021). Climate change can modify pathogen abundance and behavior, disrupt host-pathogen interactions, and stimulate the formation of novel diseases (Cohen and Leach, 2020).

Effect of waterlogging on plant microbiome

Microbes associated with root system have a significant influence on the soil environment, regulating numerous soil biochemical processes as well as plant growth and adaptive (Sun et al., 2024). Like other stressors, flooding has a direct influence on soil and root microbiome by gradually depleting O2 in soil pores which are filled with water. The shift from oxygenated to anoxic soil affects the microbial makeup from a preponderance of aerobic organisms, to a higher presence of facultative anaerobes, and eventually to the dominance of strict anaerobes. Flooding alters microbial communities in bulk and rhizo-sphere soils (Lin et al., 2011; Hamonts et al., 2013; Francioli et al., 2021). Because the bulk soil is the primary source of microorganisms recruited by plant roots in the rhizosphere (Bulgarelli et al., 2012; Bonito et al., 2014), flooding's impact on the microbial composition of the bulk soil can likewise influence the microbiome of the rhizosphere. Several research on rice plants have shown how flooding impacts the microbiome, mostly in terms of bacteria, although archaea, oomycetes, fungus, and viruses remain largely unknown. Flooding has been demonstrated to change rhizospheric and bulk soil microbial populations (Iniesta-Pallarés et al., 2023). Previous research has highlighted the impact of flooding on the rice phyllosphere microbiome, with Firmicutes (54%) and Bacillus (52.63%) being the leading species in flooded rice plants. According to Tian et al. (2015), the amount and duration of floods reduce plant microbial endophyte colonization. Under normal conditions, the microbiome profiling showed that the presence of beneficial microbial communities such as *Desulfitobacterium*), a nitrogen and carbon dioxide-fixing bacteria *Amnibacterium kyonggiense*, phosphatase and beta-glucosidase-producing bacteria, Streptomyces and Chaetomium pathogen inhibiting, and plant growth hormone-producing microbes like *Trichoderma*, *Talaromyces Promicromonospora* and Penicillium (Hyakumachi, 1994; Salas-Marina et al., 2011). However, soil microbiome profiling in sugarcane during waterlogging showed the dominance of plant detrimental microbial communities like pathogens and growth-inhibiting bacteria (Leelastwattanagul et al., 2023). Waterlogging also effects plant mycobiome in sugarcane by increasing Basidiomycota and reducing Ascomycota which contains many plant pgroth promoting fungal genera like *Trichoderma, Aspergillus, Talaromyces, Exophiala*, *Cladosporium, Phoma, Penicillium, Purpureocillium, Chaetomium*, and *Phomopsis* (Leelastwattanagul et al., 2023). Similarly, *Myricaria laxiflora*, a riparian shrub that frequently encounters periodic summer floods, has decreased endophyte diversity in anaerobic conditions. A recent study on spring wheat (*Triticum aestivum*) found that flooding stress causes substantial alterations in the makeup of the rhizosphere microbiome (Francioli et al., 2022). They found that anaerobic bacteria belonging to phyla Desulfobacterota and Firmicutes along with plant detrimental microbial taxa Geobacter and Clostridium were dominant than plant-beneficial bacterial taxa like *Sphingomonas* and *Streptomyces* which will have huge outcome on plant fitness and survival. There have been numerous studies on how flooding effected different microbial communities and their functional attributes. For instance, hypoxia triggered by flooding effects the plant mycorrhizal association mainly by inhibiting hyphal growth and AM spore germination (Tacon et al, 1983). Similarly, flooding also affects ecto-mycorrhiza (ECM) colonization and richness (Unger et al., 2009). Plants exposed to waterlogging stress reduced their ability to colonize with microbial endophytes as most of the endophytes colonizing terrestrial roots are obligate aerobes, and their survival is hindered under hypoxic conditions triggered by flooding (Li et al., 2010). Previous studies have reported that endophyte diversity was decreased in *Myricaria laxiflora* and rice plants during flooding stress (Tian et al., 2015). Flooding also alters phyllosphere microbiome structure in plants (Tian et al., 2015; Vishwanathan et al, 2020). For instance, rice culms exhibited a decrease in Gammaproteobacteria members in response to flooding stress, although Firmicutes members, particularly Bacillus species, appeared to adapt to flooding (Cui et al., 2019). In rice, flooding and heat stress dramatically changes root microbiome by enhancing the presence of bacterial alpha diversity and reducing the relative richness of Actinobacteria and Firmicutes which plays key role in carbon decomposition and soil fertility (Liu et al., 2023). In addition to root microbiome, flooding also effects leaf microbiome dynamics depending on the developmental stage, with younger plants experiencing a more dramatic disturbance in community formation (Francioli et al., 2022). Importantly, these studies reported that the change in microbiome composition was directly related to plant growth and development as well as adaptive responses.

In comparison with other stressors, how waterlogging changes plant root exudate chemistry that influence microbiome structure is not fully understood. It is well documented that plants undergo metabolic reprogramming from aerobic to anaerobic energy synthesis which can direct influence on root exudates. Therefore, it will be interesting to unravel the root exude diversity under waterlogging conditions in both model and crop plants which can provide novel insights on how plants influence its microbiome during waterlogging. However, the effect of waterlogging on soil physicochemical traits such as pH, structure, porosity, nutrients and

oxygen reduction or reduced gaseous exchange are the primary factors have severe influence on microbial diversity and community activity (Neatrour et al., 2004; Yu et al., 2022). Further we have shown the effect of waterlogging on plant microbiome and its functional traits that are associated with plant growth and adaptive responses in Figure 1. In this schematic illustration we have highlighted host driven factors like metabolic shift from aerobic to anaerobic, altered root exudes, and soil based factors such as hypoxia, reduction in gaseous exchange, nutrition shortage that alter plant microbiome during waterlogging.

Waterlogging increases pathogen distribution and disease severity in plants

Plants are constantly challenged by different microbial pathogens which causes significant yield losses (Ali et al., 2018; Mir et al., 2021). With climate change and occurrence of abiotic stressors has changed the distribution, host specificity and pathogenicity of microbial pathogens thereby posing serious threat to sustainable agriculture (Ali et al., 2023). Flooding not only affects plant growth but also makes them susceptible to number of pathogens. Flooding changes plants microbiome structure which have significant impact of their disease resistance. Indeed, flooding impacts the onset and development of various plant diseases by altering host vulnerability as well as the survival and pathogen distribution. In general, diseases and pests proliferate rapidly in high humidity circumstances due to enhanced germination and proliferation which ultimately leads to huge crop yield losses (Savary et al., 2019). Flooding promotes disease progression when plants get infected with oomycete or fungal diseases like *Phytopythium, Pythium, Phytophthora, Fusarium* (Wilcox, 1985; Moslemi et al., 2018). These pathogens cause more damage to waterlogged stress plants and leads to high mortality. During floods, increased root exudation of ethanol, carbohydrates, and amino acids can promote pathogen infection (Blaker and McDonald, 1981; Tyler, 2002). Flooding promotes disease development by altering beneficial microbial communities which are crucial for activating plant immune system. Previous studies have shown that flood affects immune modulator beneficial microbial communities belonging *Sphingomonas*, *Streptomyces, Flavobacterium, Saccharimonadia* and *Massilia* which leads to disease progression (Soltani et al., 2010; Kavamura et al., 2021). Also, flooding promotes disease progression by affecting plant symbiotic association which are known to inhibit disease and pathogen distribution by their antagonistic or antibiotic potential, as well as activating systemic resistance induction. Francioli et al. (2021), reported an increase in *Clostridium* species in roots after floods which is commonly linked with root rot under waterlogged soils. Previous study has revealed that flooding in Ulmus minor plants changes root mycobiome and increases the development of root rot disease caused by *Plectosphaerella cucumerina* by altering the beneficial microbial communities (Martínez-Arias et al., 2020). Recent study has revealed that *Phytophthora medicaginis* a causative agent of root rot disease in chickpea, was more severe under waterlogging conditions (Dron et al., 2022). At present there are not effective control measures against Phytophthora root rot however, farmers are advised to avoid fields prone to waterlogging. Similar studies have demonstrated that waterlogging enhances the infection of phytopthora root rot in oak, avocado, and lucerne plants (Kuan and Erwin, 1980; Jacobs et al., 1997). In kiwi fruit, waterlogging triggers the severity of root rot disease caused by

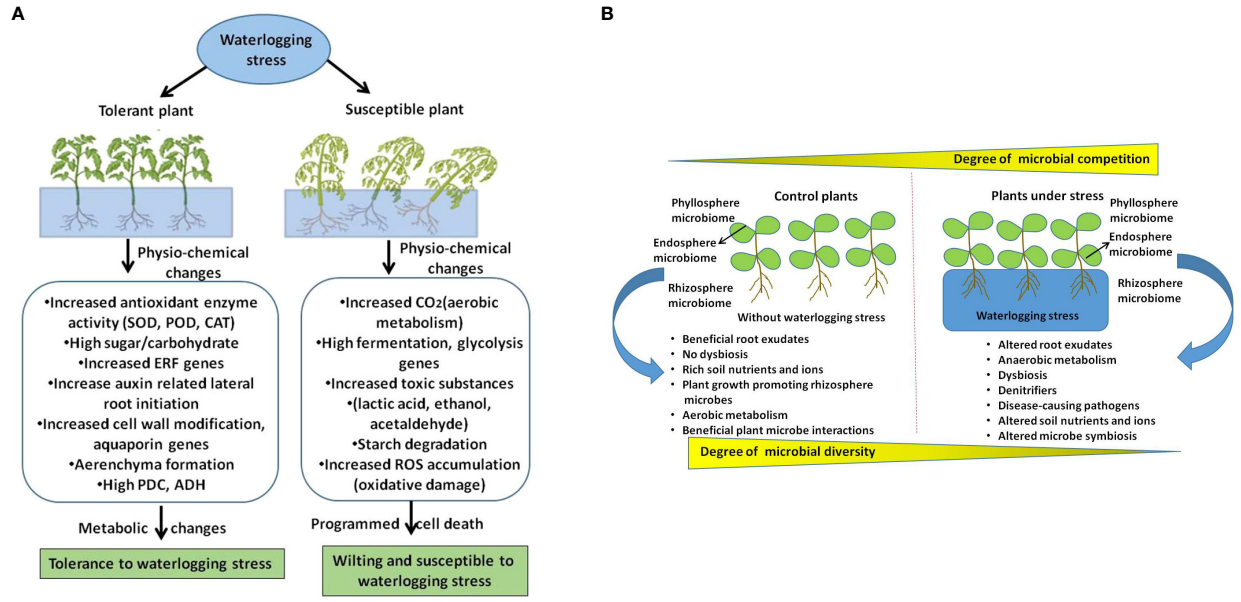


FIGURE 1
A schematic illustration shows how waterlogging affect plants and their microbiome. **(A)** Depicts waterlogging induced changes in tolerant and susceptible plants. **(B)** Shows how waterlogging stress affects microbiome structure by altering host driven factors and soil physicochemical properties which leads to dysbiosis and effects plant growth and adaptive traits.

Phytopythium vexans and *Phytopythium chamaehyphon*, which causes more damage that waterlogging alone (Savian et al., 2020). Waterlogging has been shown to increase the prevalence of apple crown and root rot (Phytophthora spp.), banana vascular wilt (*F. oxysporum* f. sp. cubense), raspberry damping off (*Pythium irregulare*), and chili pepper verticillium wilt (*Verticillium dahlia*) (Aguilar et al., 2000; Sanogo et al., 2008; Li et al., 2015). On the other hand, waterlogging renders pigeon pea plants more vulnerable to fungal diseases including Fusarium wilt and Phytophthora blight, resulting in considerable output losses (Yohan et al., 2017). Previous study has reported that waterlogging increases the severity of disease in pea plants caused by *Mycosphaerella pinodes* and resulted in reduced root and shoot growth (McDonald and Dean, 1996). Above studies further provides the evidence that waterlogging enhances pathogen aggressiveness, their occurrence that causes more damage than waterlogging alone. Further we have summarized the case studies highlighting the effect of waterlogging on disease incidence and severity in different plants in (Table 1).

The promotion of pathogenic microbes and disease progression in plants during waterlogging in mainly linked with energy deprivation (Moslemi et al., 2018), suppression of oxidative burst and the immune response, and hypersensitive cell death (McDonald, 2002). It will be interesting in future to decipher the molecular complexity of waterlogging and plant disease development and identify potential targets that suppresses plant immunity. Further we have shown the effect of waterlogging on disease development in plants in Figure 2. In contrast, hypoxia caused by floods can minimize disease development in plants by enhancing host defense responses as a result of the activation of a general stress response (Chung and Lee, 2020). Similarly, it was found that flooding stress lead the activation of plant immune signature transcriptional factors WRKY which modulated the expression of plant defense marker genes which have diverse antagonistic effect on microbial pathogens (Hsu et al., 2013). Future research is thus needed to determine how waterlogging affects plant immune signaling cascades utilizing different crop systems, since this would offer fresh perspectives on enhancing disease resistance.

TABLE 1 Waterlogging induced pathogen occurrence and disease progression in different plants.

Plant	Waterlogging (WL)	Pathogen	Effect	References
Chickpea	WL	*Phytophthora medicaginis*	Enhanced disease progression and reduced the total plant biomass	(Dron et al., 2022)
Soybean	WL	*Pythium ultimum P. irregulare, P. aphanidermatum, and P. vexans*	Root discoloration. Reduced plant weight	(Kirkpatrick et al., 2006)
Wheat and barley	WL	*Fusarium poae*	Enhanced disease incidence and severity, Effects yield components and grain composition	(Martínez et al., 2019)
Bell Pepper	WL	*Phytophthora capsici*	Increased disease severity and plant mortality	(Bowers et al., 1990)
Apple	WL	*Phytophthora cambivora, Phytophthora cactorum* and *P. cryptogea*	Enhanced the severity of root rot disease	(Browne and Mircetich, 1988)
Soybean and tobacco	WL	*P. sojae and P. nicotianae*	Suppression of host disease resistance	(McDonald, 2002)
Kiwifruit	WL	*P. vexans and P. chamaehyphon*	Enhanced onset of disease and severity. High mortality rate	(Savian et al., 2020)
Potato	WL	*Erwinia carotovora*	Effect the expression of host defense genes and increased the severity soft rot.	(Rumeau et al., 1990)
Onion	WL	*Colletotrichum* sp.	Enhanced disease development	(Lopes et al., 2021)
Banana	WL	*F. oxysporum*	Increased the prevalence of wilt disease	(Aguilar et al., 2000)
Common bean	WL	*Pythium sps*	Enhanced disease development	(Li et al., 2015)
Pigeon pea	WL	*Fusarium and Phytophthora*	Increased disease incidence and development. Yield losses	(Yohan et al., 2017)
Peach	WL	*P. vexans and P. irregulare*	Enhanced disease severity	(Biesbrock and Hendrix, 1970)
Pea	WL	*Mycosphaerella pinodes*	Increased disease incidence Reduced root and shoot growth	(McDonald and Dean, 1996)
Alfalfa	WL	*Phytophthora megasperma*	Enhanced root damage and disease severity	(Kuan and Erwin, 1980)
Jarrah	WL	*Phytophthora cinnamomi*	Increased disease incidence and severity	(Burgess et al., 1999)

10.3389/fpls.2024.1407789

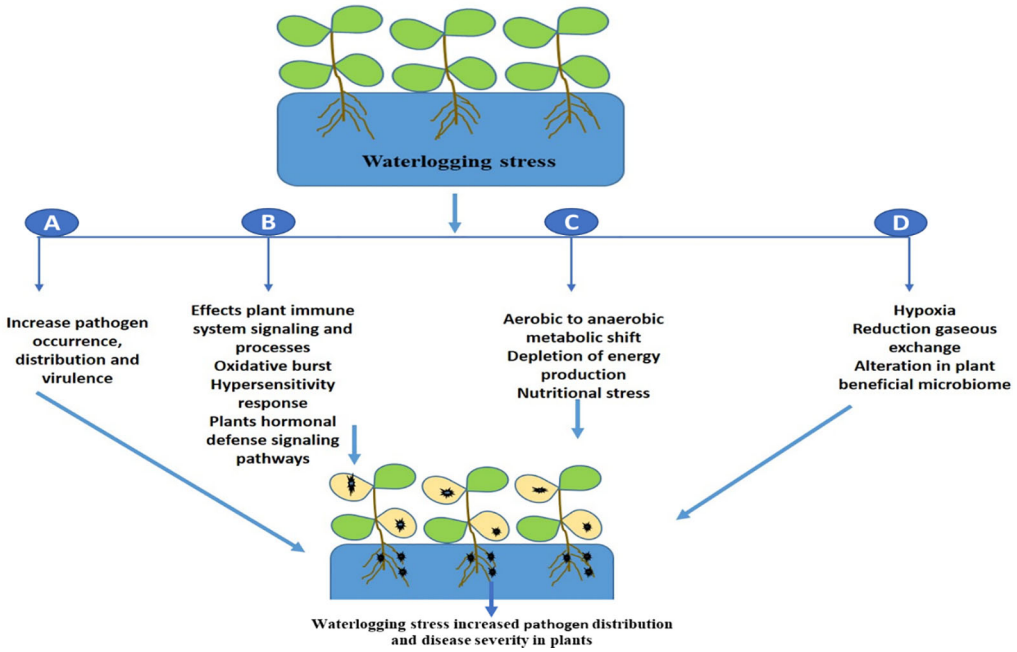


FIGURE 2
A schematic illustration shows how waterlogging affects different plant traits that promotes pathogen distribution and disease development. **(A)** waterlogging can enhance pathogen invasion or virulence, **(B)** it affects plant immunity signaling cascades which leads plant more vulnerable to diseases, **(C)** waterlogging induced metabolic shift, hypoxia, energy crisis and nutritional stress also promotes disease progression, **(D)** Alterations in beneficial microbiome or dysbiosis leads to pathogen dominance and disease progression.

Conclusion and future perspectives

Waterlogging is a complex process which affects plant growth and its metabolic traits by reducing soil oxygen levels, soil nutrient utilization efficiency and altering microbiome. Over the last decade, there has been a great deal of interest in exploring the beneficial functional attributes of plant microbiomes, for crop improvement which have proven to be an effective nature-based solution to combat environmental stresses while safeguarding environmental and soil health. For instance, the development of drought or salinity tolerant microbiome synComs have been used by many researchers to improve the growth and adaptive responses in different plant systems under lab and field conditions. Therefore, there is need to utilize plant beneficial microbiome as a key strategy for mitigating waterlogging stress in plants and enhance crop productivity. This will require an in-depth study using different omics and synthetic biology approaches to decipher how plant microbiome responds and adapts to waterlogging stress. Also, harnessing genome editing tools and synthetic biology to engineer plants to produce root exudates that can shape unique stress resilient microbiome is another viable strategy for improving plant waterloging tolerance and growth traits. There is need to explore the microbiome of waterlogging tolerant crops or wild varieties which may lead to the identification unique stress resilient microbiota that might promote the plant growth and survival under waterlogging conditions. Future studies should also unwire how waterlogging driven metabolic shift in plants alter root exudate chemistry that will eventually shape distinct microbiome communities. In light of climate change, microbes are the best candidates to explore because of their rapid natural adaptability to environmental extremes and nourish under extreme growth conditions. At the same time, how microbiome dysbiosis occurs during waterlogging stress that triggers disease development or pathogen distribution warrants future investigation. Finally, we recomend for designing future flood modelling tools which will be an effective way for evaluating hypothesis and examining different situations, particularly with relation to the health of plants, microbial communities, plant pathogen interactions and soil health. This would require the collaboration of plant scientists from different backgrounds to design future flood modelling in order to prevent flood induced damage to sustainable agriculture.

Author contributions

AT: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration, Funding acquisition. SA: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration, Funding acquisition. RAM: Methodology, Software, Validation, Formal analysis, Data curation, Writing – original draft, Writing –

review & editing. SS: Software, Writing – review & editing, Visualization. KA: Software, Writing – review & editing. MA: Conceptualization, Writing – original draft, Writing – review & editing, Supervision. ZAM: Methodology, Software, Validation, Formal analysis, Data curation, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, 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.

Publisher's note

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

References

Aguilar, E. A., Turner, D. W., and Sivasithamparam, K. (2000). *Fusarium oxysporum* f. sp. cubense inoculation and hypoxia alter peroxidase and phenylalanine ammonia-lyase activities in nodal roots of banana cultivars (Musa sp.) differing in their susceptibility to Fusarium wilt. *Aust. J. Bot.* 48, 589–596. doi: 10.1071/BT99009

Ali, S., Ganai, B. A., Kamili, A. N., Bhat, A. A., Mir, Z. A., Bhat, J. A., et al. (2018). Pathogenesis-related proteins and peptides as promising tools for engineering plants with multiple stress tolerance. *Microbiol. Res.* 212, 29–37. doi: 10.1016/j.micres.2018.04.008

Ali, S., and Kim, W. C. (2018). Plant growth promotion under water: decrease of waterlogging-induced ACC and ethylene levels by ACC deaminase-producing bacteria. *Front. Microbiol.* 9. doi: 10.3389/fmicb.2018.01096

Ali, S., Tyagi, A., and Bae, H. (2023). Plant microbiome: an ocean of possibilities for improving disease resistance in plants. *Microorganisms.* 11, 392. doi: 10.3390/microorganisms11020392

Ali, S., Tyagi, A., Mushtaq, M., Al-Mahmoudi, H., and Bae, H. (2022a). Harnessing plant mi-crobiome for mitigating arsenic toxicity in sustainable agriculture. *Environ. pollut.* 300, 118940. doi: 10.1016/j.envpol.2022.118940

Ali, S., Tyagi, A., Park, S., Mir, R. A., Mushtaq, M., Bhat, B., et al. (2022b). Deciphering the plant microbiome to improve drought tolerance: mechanisms and perspectives. *Environ. Exp. Bot.* 201, 104933. doi: 10.1016/j.envexpbot.2022.104933

Ashraf, M. A. (2012). Waterlogging stress in plants: A review. *Afr. J. Agric. Res.* 7, 1976–1981. doi: 10.5897/AJARX11.084

Biesbrock, J. A., and Hendrix, F. F. (1970). Influence of soil water and temperature on root necrosis of peach caused by Pythium spp. *Phytopathol.* 60, 880–882. doi: 10.1094/Phyto-60-880

Blaker, N. S., and McDonald, J. D. (1981). Predisposing effects of soil moisture extremes on the susceptibility of Rhododendron to Phytophthora Root and crown rot. *Phytopathol.* 71, 831. doi: 10.1094/Phyto-71-831

Bokhari, A., Essack, M., Lafi, F. F., Andres-Barrao, C., Jalal, R., Alamoudi, S., et al (2019). Bioprospecting desert plant Bacillus endophytic strains for their potential to enhance plant stress tolerance. *Sci. Rep.* 9, 1–13. doi: 10.1038/s41598-019-54685-y

Bonito, G., Reynolds, H., Robeson, M. S., Nelson, J., and Hodkinson, B. P. (2014). Plant host and soil origin influence fungal and bacterial assemblages in the roots of woody plants. *Mol. Ecol.* 23, 3356–3370. doi: 10.1111/mec.12821

Bowers, J. H., Sonoda, R. M., and Mitchell, D. J. (1990). Path coefficient analysis of the effect of rainfall variables on the epidemiology of Phytophthora blight of pepper caused by Phytophthora capsici. *Phytopathol.* 80, 1439–1446. doi: 10.1094/Phyto-80-1439

Browne, G. T., and Mircetich, S. M. (1988). Effects of flood duration on the development of Phytophthora root and crown rots of apple. *Phytopathol.* 78, 846:851. doi: 10.1094/Phyto-78-846

Bulgarelli, D., Rott, M., Schlaeppi, K., Ver Loren van Themaat, E., Ahmadinejad, N., Assenza, F., et al. (2012). Revealing structure and assembly cues for Arabidopsis root-inhabiting bacterial microbiota. *Nat.* 488, 91–95. doi: 10.1038/nature11336

Burgess, T., McComb, J. A., Colquhoun, I., and Hardy, G. E. S. S. (1999). Increased susceptibility of Eucalyptus marginata to stem infection by Phytophthora cinnamomi resulting from root hypoxia. *Plant Patho.* 48, 797–806. doi: 10.1046/j.1365-3059.1999.00396.x

Cheng, X. F., Wu, H. H., Zou, Y. N., Wu, Q. S., and Kuča, K. (2021). Mycorrhizal response strategies of trifoliate orange under well-watered, salt stress, and waterlogging stress by regulating leaf aquaporin expression. *Plant Physiol. Biochem.* 162, 27–35. doi: 10.1016/j.plaphy.2021.02.026

Chung, H., and Lee, Y. H. (2020). Hypoxia: a double-edged sword during fungal pathogenesis? *Front. Microbiol.* 11. doi: 10.3389/fmicb.2020.01920

Cohen, S. P., and Leach, J. E. (2020). High temperature-induced plant disease susceptibility: more than the sum of its parts. *Curr. Opin. Plant Biol.* 56, 235–241. doi: 10.1016/j.pbi.2020.02.008

Colmer, T. D., Kotula, L., Malik, A. I., Takahashi, H., Konnerup, D., Nakazono, M., et al. (2019). Rice acclimation to soil flooding: low concentrations of organic ac-ids can trigger a barrier to radial oxygen loss in roots. *Plant Cell Environ.* 42, 2183–2197. doi: 10.1111/pce.13562

Compant, S., Samad, A., Faist, H., and Sessitsch, A. (2019). A review on the plant microbiome: ecology, functions, and emerging trends in microbial application. *J. Adv. Res.* 19, 29–37. doi: 10.1016/j.jare.2019.03.004

Cui, H. L., Duan, G. L., Zhang, H., Cheng, W., and Zhu, Y. G. (2019). Microbiota in non-flooded and flooded rice culms. *FEMS Microbiol. Ecol.* 95, fiz036. doi: 10.1093/femsec/fiz036

Dron, N., Merrill Ryan, W. M., Forknall, C., Hobson, K., Sutton, T., and Bithell, S. (2022). Phytophthora root rot and waterlogging in chickpeas–minimising risk and man-agement options. *Grains Res. Update* 100.

FAO. (2021). *The state of food security and nutrition in the world. Trans-forming food systems for food security, improved nutrition and affordable healthy diets for all* (Rome: Food and Agriculture Organization of the United Nations). Available at: https://www.fao.org/documents/card/en?details=cb4476en.

Farwell, A. J., Vesely, S., Nero, V., Rodriguez, H., McCormack, K., Shah, S., et al. (2007). Tolerance of transgenic canola plants (Brassica napus) amended with plant growth-promoting bacteria to flooding stress at a metal-contaminated field site. *Environ. pollut.* 147, 540–545. doi: 10.1016/j.envpol.2006.10.014

Fitzpatrick, C. R., Salas-González, I., Conway, J. M., Finkel, O. M., Gilbert, S., Russ, D., et al. (2020). The plant microbiome: from ecology to reductionism and beyond. *Annu. Rev. Microbiol.* 74, 81–100. doi: 10.1146/annurev-micro-022620-014327

Francioli, D., Cid, G., Hajirezaei, M. R., and Kolb, S. (2022). Leaf bacterial microbiota response to flooding is controlled by plant phenology in wheat (Triticum aestivum L.). *Sci. Rep.* 12 (1), 11197. doi: 10.1038/s41598-022-15133-6

Francioli, D., Cid, G., Kanukollu, S., Ulrich, A., Hajirezaei, M.-R., and Kolb, S. (2021). Flooding causes dramatic compositional shifts and depletion of putative beneficial bacteria on the Spring wheat microbiota. *Front. Microbiol.* 12. doi: 10.3389/fmicb.2021.773116

Furtak, K., and Wolińska, A. (2023). The impact of extreme weather events as a consequence of climate change on the soil moisture and on the quality of the soil environment and agriculture–A review. *Catena* 231, 107378. doi: 10.1016/j.catena.2023.107378

Hamonts, K., Clough, T. J., Stewart, A., Clinton, P. W., Richardson, A. E., Wakelin, S. A., et al. (2013). Effect of nitrogen and waterlogging on denitri-fier gene abundance, community structure and activity in the rhizosphere of wheat. *FEMS Microbiol. Ecol.* 83, 568–584. doi: 10.1111/fem.2013.83.issue-3

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

Hsu, F. C., Chou, M. Y., Peng, H. P., Chou, S. J., and Shih, M. C. (2011). Insights into hypoxic systemic responses based on analyses of transcriptional regulation in Arabidopsis. *PloS One* 6, e28888. doi: 10.1371/journal.pone.0028888

Hsu, F.-C., and Shih, M.-C. (2013). Plant defense after flooding. *Plant Signaling Behav.* 8, 2699–2713. doi: 10.4161/psb.26922

Hubbard, C. J., Li, B., McMinn, R., Brock, M. T., Maignien, L., Ewers, B. E., et al. (2019). The effect of rhizosphere microbes outweighs host plant genetics in re-ducing insect herbivory. *Mol. Ecol.* 28, 1801–1811. doi: 10.1111/mec.14989

Hyakumachi, M. (1994). Plant-growth-promoting fungi from turfgrass rhizosphere with potential for disease suppression. *Soil Microorganisms.* 44, 53–68. doi: 10.18946/jssm.44.0_53

Iniesta-Pallarés, M., Brenes-Álvarez, M., Lasa, A. V., Fernández-López, M., Álvarez, C., Molina-Heredia, F. P., et al. (2023). Changes in rice rhizosphere and bulk soil bacterial communities in the Doñana wetlands at different growth stages. *Appl. Soil Ecol.* 190, 105013. doi: 10.1016/j.apsoil.2023.105013

Jacobs, K. A., MacDonald, J. D., Berry, A. M., and Costello, L. R. (1997). *Rooting responses of three oak species to low oxygen stress, proceedings of a symposium on oak woodlands: Ecology, management and urban interface issues* (Albany, CA: Gen. Tech. Rep. USDA Forest Service), 91–100.

Jia, R., Ma, X., and Xie, V. W. (2022). *Expecting floods: Firm entry, employment, and aggregate implications (No. w30250)* (National Bureau of Economic Research). doi: 10.3386/w30250

Jongman, B., Koks, E. E., Husby, T. G., and Ward, P. J. (2014). Increasing flood exposure in the Netherlands: implications for risk financing. *NHESS.* 14, 1245–1255. doi: 10.5194/nhess-14-1245-2014

Kavamura, V. N., Mendes, R., Bargaz, A., and Mauchline, T. H. (2021). Defining the wheat microbiome: towards microbiome-facilitated crop production. *CSBJ* 19, 1200–1213.c. doi: 10.1016/j.csbj.2021.01.045

Kirkpatrick, M. T., Rupe, J. C., and Rothrock, C. S. (2006). Soybean response to flooded soil conditions and the association with soilborne plant pathogenic genera. *Plant Dis.* 90, 592–596. doi: 10.1094/PD-90-0592

Koskella, B., Meaden, S., Crowther, W. J., Leimu, R., and Metcalf, C. J. E. (2017). A signature of tree health? Shifts in the microbiome and the ecological drivers of horse chestnut bleeding canker disease. *New Phytol.* 215, 737–746. doi: 10.1111/nph.14560

Kreibich, H., Van Loon, A. F., Schröter, K., Ward, P. J., Mazzoleni, M., Sairam, N., et al. (2022). The challenge of unprecedented floods and droughts in risk management. *Nature* 608, 80–86. doi: 10.1038/s41586-022-04917-5

Kron, W. (2005). Flood risk hazard values vulnerability. *Water Int.* 30, 58–68. doi: 10.1080/02508060508691837

Kuan, T. L., and Erwin, D. C. (1980). Predisposition effect of water saturation of soil on Phytophthora root rot of Alfalfa. *Phytopathol.* 70, 981–986. doi: 10.1094/Phyto-70-981

Lau, J. A., Lennon, J. T., and Heath, K. D. (2017). Trees harness the power of microbes to survive climate change. *PNAS.* 114, 11009–11011. doi: 10.1073/pnas.1715417114

Leelastwattanagul, O., Sutheeworapong, S., Khoiri, A. N., Dulsawat, S., Wattanachaisaereekul, S., Tachaleat, A., et al. (2023). Soil microbiome analysis re-veals effects of periodic waterlogging stress on sugarcane growth. *PloS One* 18, e0293834. doi: 10.1371/journal.pone.0293834

Li, H. Y., Zhao, C. A., Liu, C. J., and Xu, X.-F. (2010). Endophytic fungi diversity of aquatic/riparian plants and their antifungal activity in *vitro*. *J. Microbiol.* 48, 1–6. doi: 10.1007/s12275-009-0163-1

Li, M., López, R., Venturas, M., Pita, P., Gordaliza, G. G., Gil, L., et al. (2015). Greater resistance to flooding of seedlings of Ulmus laevis than Ulmus minor is related to the maintenance of a more positive carbon balance. *Trees.* 29, 835–848. doi: 10.1007/s00468-015-1163-x

Lin, X., Zhu, D., and Lin, X. (2011). Effects of water management and organic fertilization with SRI crop practices on hybrid rice performance and rhizosphere dynamics. *PAWE.* 9, 33–39. doi: 10.1007/s10333-010-0238-y

Liu, B., Sun, L., Ma, L., and Hao, F. S. (2017). Both AtrbohD and AtrbohF are essential for mediating responses to oxygen deficiency in Arabidopsis. *Plant Cell Rep.* 36, 947–957. doi: 10.1007/s00299-017-2128-x

Liu, J., Zhen, B., Qiu, H., Zhou, X., and Zhang, H. (2023). Impact of waterlogging and heat stress on rice rhizosphere microbiome assembly and potential function in carbon and nitrogen transformation. *Arch. Agron. Soil Sci.* 69, 1920–1932. doi: 10.1080/03650340.2022.2128190

Lopes, L. H. R., Boiteux, L. S., Rossato, M., Aguiar, F. M., Fonseca, M. E., Oliveira, V. R., et al. (2021). Diversity of Colletotrichum species causing onion anthracnose in Brazil. *Eur. J. Plant Pathol.* 159, 339–357. doi: 10.1007/s10658-020-02166-8

Martínez, M., Arata, A. F., Lázaro, L., Stenglein, S. A., and Dinolfo, M. I. (2019). Effects of waterlogging stress on plant-pathogen interaction between Fusarium poae and wheat/barley. Acta Scientiarum. *Agron.* 41, e42629. doi: 10.4025/actasciagron.v41i1.42629

Martínez-Arias, C., Sobrino-Plata, J., MacAya-Sanz, D., Aguirre, N. M., and Collada, C. (2020). Changes in plant function and root mycobiome caused by flood and drought in a riparian tree. *Tree Physiol.* 40, 886–903. doi: 10.1093/treephys/tpaa031

McDonald, K. (2002). Temporary hypoxia suppresses the oxidative burst and subsequent hypersensitive cell death in cells of tobacco and soybean challenged with zoospores of incompatible isolates of Phytophthora species. *Physiol. Plant Pathol.* 61, 133–140. doi: 10.1006/pmpp.2002.0423

McDonald, G. K., and Dean, G. (1996). Effect of waterlogging on the severity of disease caused by Mycosphaerella pinodes in peas (Pisum sativum L.). *Aust. J. Exp. Agric.* 36, 219–222. doi: 10.1071/EA9960219

Mir, Z. A., Ali, S., Singh, A., Yadav, P., Tyagi, A., Chaturani, G. D. G., et al. (2021). In silico analysis and overexpression of chitinase class IV gene in Brassica juncea improves resistance against Alternaria brassicae. *Indus. Crops Prod.* 169, 113555. doi: 10.1016/j.indcrop.2021.113555

Moslemi, A., Ades, P. K., Groom, T., Nicolas, M. E., and Taylor, P. W. (2018). Influence of waterlogging on growth of pyrethrum plants infected by the crown and root rot pathogens, Fusarium oxysporum, Fusarium avenaceum and Paraphoma vinacea. *Australas. Plant Pathol.* 47, 205–213. doi: 10.1007/s13313-018-0547-y

Neatrour, M. A., Webster, J. R., and Benfield, E. F. (2004). The role of floods in particulate organic matter dynamics of a southern Appalachian river–floodplain ecosystem. *J. North Am. Benthological Society.* 23, 198–213. doi: 10.1899/0887-3593(2004)023<0198:TROFIP>2.0.CO;2

Nio, S. A., and Mantilen, D. P. (2023). Beneficial root-associated microbiome during drought and flooding stress in plants. *Pak J. Biol. Sci.* 26, 287–299. doi: 10.3923/pjbs.2023.287.299

Peláez-Vico, M.Á., Tukuli, A., Singh, P., Mendoza-Cózatl, D. G., Joshi, T., and Mittler, R. (2023). Rapid systemic responses of Arabidopsis to waterlogging stress. *Plant Physiol.* 193, 2215–2231. doi: 10.1093/plphys/kiad433

Rumeau, D., Maher, E. A., Kelman, A., and Showalter, A. M. (1990). Extensin and Phenylalanine Ammonia-Lyase gene expression altered in potato tubers in response to wounding, hypoxia, and Erwinia carotovora infection. *Plant Physiol.* 93, 1134–1139. doi: 10.1104/pp.93.3.1134

Salas-Marina, M. A., Silva-Flores, M. A., Cervantes-Badillo, M. G., Rosales-Saavedra, M. T., Islas-Osuna, M. A., and Casas-Flores, S. (2011). The plant growth-promoting fungus Aspergillus ustus promotes growth and induces resistance against different lifestyle pathogens in Arabidopsis thaliana. *JMB.* 21, 686–696. doi: 10.4014/jmb.1101.01012

Sanogo, S., El-Sebai, O. I., and Sanderson, R. (2008). Severity of Verticillium wilt, plant growth, and spectral reflectance indices of Chile pepper under periodic flooding and no-flooding conditions. *Hort Sci.* 43, 414–419. doi: 10.21273/HORTSCI.43.2.414

Sasidharan, R., Bailey-Serres, J., Ashikari, M., Atwell, B. J., Colmer, T. D., Fagerstedt, K., et al. (2017). Community recommendations on terminology and procedures used in flooding and low oxygen stress research. *New Phyt.* 214, 1403–1407. doi: 10.1111/nph.14519

Sauter, M. (2013). Root responses to flooding. *Curr. Opin. Plant Biol.* 16, 282–286. doi: 10.1016/j.pbi.2013.03.013

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

Savian, F., Ginaldi, F., Musetti, R., Sandrin, N., Tarquini, G., Pagliari, L., et al. (2020). Studies on the aetiology of kiwifruit decline: interaction between soil-borne pathogens and waterlogging. *Plant Soil.* 456, 113–128. doi: 10.1007/s11104-020-04671-5

Shekhawat, K., Saad, M. M., Sheikh, A., Mariappan, K., Al-Mahmoudi, H., Abdulhakim, F., et al. (2021). Root endophyte induced plant thermotolerance by constitutive chromatin modification at heat stress memory gene loci. *EMBO Rep.* 22, e51049. doi: 10.15252/embr.202051049

Shukla, V., Lombardi, L., Iacopino, S., Pencik, A., Novak, O., Perata, P., et al. (2019). Endogenous hypoxia in lateral root primordia controls root architec-ture by antagonizing auxin signaling in Arabidopsis. *Mol. Plant* 12, 538–551. doi: 10.1016/j.molp.2019.01.007

Soltani, A.-A., Khavazi, K., Asadi-Rahmani, H., Omidvari, M., Abaszadeh Da-haji, P., and Mirhoseyni, H. (2010). Plant growth promoting characteristics in some Flavobacterium spp. isolated from soils of Iran. *J. Agric. Sci.* 2, 106. doi: 10.5539/jas.v2n4p106

Sun, C., Lu, X., Wang, Y., and Qiu, M. (2024). Regulation of soil nutrient cycling in the root zone of Pyracantha fortuneana: The role of core microbiome induced by plant species. *JPNSS.* doi: 10.1002/jpln.202300372

Tacon, F. L., Skinner, F. A., and Mosse, B. (1983). Spore germination and hyphal growth of a vesicular–arbuscular mycorrhizal fungus, Glomus mosseae (Gerdemann and Trappe), under decreased oxygen and increased carbon dioxide concentrations. *Can. J. Microbiol.* 29 (10), 1280–1285.

Tamang, B. G., Magliozzi, J. O., Maroof, M. A., and Fukao, T. (2014). Physiological and tran-scriptomic characterization of submergence and reoxygenation responses in soybean seedlings. *Plant Cell Environ.* 37, 2350–2365. doi: 10.1111/pce.12277

Tian, B. Y., Cao, Y., and Zhang, K. Q. (2015). Metagenomic insights into communities, functions of endophytes and their associates with infection by root-knot nematode, Meloidogyne incognita, in tomato roots. *Sci. Rep.* 5, 17087. doi: 10.1038/srep17087

Timmusk, S., Pall, T., Raz, S., Fetsiukh, A., and Nevo, E. (2023). The potential for plant growth-promoting bacteria to impact crop productivity in future agricultural systems is linked to understanding the principles of microbial ecology. *Front. Microbiol.* 14. doi: 10.3389/fmicb.2023.1141862

Tonkin, J. D., Merritt, D. M., Olden, J. D., Reynolds, L. V., and Lytle, D. A. (2018). Flow regime alteration degrades ecological networks in riparian ecosystems. *Nat. Ecol. Evol.* 2, 86–93. doi: 10.1038/s41559-017-0379-0

Trivedi, P., Leach, J. E., Tringe, S. G., Sa, T., and Singh, B. K. (2020). Plant–microbiome interactions: from community assembly to plant health. *Nat. Rev. Microbiol.* 18, 607–621. doi: 10.1038/s41579-020-0412-1

TSAI, K. J., CHOU, S. J., and SHIH, M. C. (2014). Ethylene plays an essential role in the recovery of Arabidopsis during post-anaerobiosis reoxygenation. *Plant Cell Environ.* 37, 2391–2405. doi: 10.1111/pce.12292

Tyagi, A., Sharma, S., Srivastava, H., Singh, A., Kaila, T., Ali, S., et al. (2023). Transcriptome profiling of two contrasting pigeon pea (*Cajanus cajan*) geno-types in response to waterlogging stress. *Front. Genet.* 13. doi: 10.3389/fgene.2022.1048476

Tyler, B. M. (2002). Molecular basis of recognition between Phytophthora pathogens and their hosts. *Annu. Rev. Phytopathol.* 40, 137–167. doi: 10.1146/annurev.phyto.40.120601.125310

Unger, I. M., Kennedy, A. C., and Muzika, R. M. (2009). Flooding effects on soil microbial communities. *Appl. Soil Ecol.* 42, 1–8. doi: 10.1016/j.apsoil.2009.01.007

Vishwanathan, K., Zienkiewicz, K., Liu, Y., Janz, D., Feussner, I., Polle, A., et al. (2020). Ectomycorrhizal fungi induce systemic resistance against insects on a nonmycorrhizal plant in a CERK1-dependent manner. *New Phytol.* 228, 728–740. doi: 10.1111/nph.16715

Voesenek, L., and Bailey-Serres, J. (2015). Flood adaptive traits and processes: an overview. *New Phytol.* 206, 57–73. doi: 10.1111/nph.13209

Voesenek, L. A., Sasidharan, R., Visser, E. J., and Bailey-Serres, J. (2016). Flooding stress signaling through perturbations in oxygen, ethylene, nitric oxide and light. *New Phytol.* 209, 39–43. doi: 10.1111/nph.13775

Wilcox, W. F. (1985). Effects of flooding duration on the development of Phytophthora root and crown rots of cherry. *Phytopathol.* 75, 1451. doi: 10.1094/Phyto-75-1451

Wille, L., Messmer, M. M., Studer, B., and Hohmann, P. (2019). Insights to plant–microbe interactions provide opportunities to improve resistance breeding against root diseases in grain legumes. *Plant Cell Environ.* 42, 20–40. doi: 10.1111/pce.13214

Wipf, H. M. L., Bùi, T. N., and Coleman-Derr, D. (2021). Distinguishing between the impacts of heat and drought stress on the root microbiome of Sorghum bicolor. *Phytobiomes J.* 5, 166–176. doi: 10.1094/PBIOMES-07-20-0052-R

Xu, Y., Zhang, G., Ding, H., Ci, D., Dai, L., and Zhang, Z. (2020). Influence of salt stress on the rhizosphere soil bacterial community structure and growth performance of groundnut (Arachis hypogaea L.). *Int. J. Microbiol.* 23, 453–465. doi: 10.1007/s10123-020-00118-0

Yang, L., Li, N., Liu, Y., Miao, P., Liu, J., and Wang, Z. (2023). Updates and prospects: morphological, physiological, and molecular regulation in crop response to waterlogging stress. *Agron.* 13. doi: 10.3390/agronomy13102599

Yang, J., Mathew, I. E., Rhein, H., Barker, R., Guo, Q., Brunello, L., et al. (2022). The vacuolar H+/Ca transporter CAX1 participates in sub-mergence and anoxia stress responses. *Plant Physiol.* 190, 2617–2636. doi: 10.1093/plphys/kiac375

Yohan, Y., Umamahesh, V., Sudhakar, P., Reddiramu, Y., and Sreedhar, D. (2017). Influence of waterlogging on certain biochemical and yield parameters of pigeon-pea (Cajanus cajan (L.) millsp). *Int. J. Pure App. Biosci.* 5, 1862–1868. doi: 10.18782/2320-7051.2964

Yu, T., Cheng, L., Liu, Q., Wang, S., Zhou, Y., Zhong, H., et al. (2022). Ef-fects of waterlogging on soybean rhizosphere bacterial community using V4, LoopSeq, and PacBio 16S rRNA Sequence. *Microbiol. Spectr.* 10, e02011–e02021. doi: 10.1128/spectrum.02011-21

Zhou, X., Zhang, J., Gao, D., Wei, Z., Wu, F., and Dini-Andreote, F. (2023). Interspecific plant interaction via root exudates structures the disease suppres-siveness of rhizosphere microbiomes. *Mol. Plant* 16, 849–864. doi: 10.1016/j.molp.2023.03.009

Zölch, T., Henze, L., Keilholz, P., and Pauleit, S. (2017). Regulating urban surface runoff through nature-based solutions–an assessment at the micro-scale. *Environ. Res.* 157, 135–144. doi: 10.1016/j.envres.2017.05.023

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 20 June 2024
DOI 10.3389/fpls.2024.1398437

Check for updates

OPEN ACCESS

EDITED BY
Reyazul Rouf Mir,
Sher-e-Kashmir University of Agricultural Sciences and Technology, India

REVIEWED BY
Rita Sharma,
Birla Institute of Technology and Science, Pilani, India
Veerendra Sharma ,
Kansas State University, United States

*CORRESPONDENCE
Basavaprabhu L. Patil
blpatil2046@gmail.com;
Basavaprabhu.Patil@icar.gov.in

RECEIVED 09 March 2024
ACCEPTED 29 May 2024
PUBLISHED 20 June 2024

CITATION
Patil BL and Tripathi S (2024) Differential expression of microRNAs in response to *Papaya ringspot virus* infection in differentially responding genotypes of papaya (*Carica papaya* L.) and its wild relative. *Front. Plant Sci.* 15:1398437. doi: 10.3389/fpls.2024.1398437

COPYRIGHT
© 2024 Patil and Tripathi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Differential expression of microRNAs in response to *Papaya ringspot virus* infection in differentially responding genotypes of papaya (*Carica papaya* L.) and its wild relative

Basavaprabhu L. Patil[1]* and Savarni Tripathi[2]

[1]ICAR-Indian Institute of Horticultural Research, Bengaluru, India, [2]ICAR-Indian Agricultural Research Institute, Regional Station, Pune, India

Papaya ringspot virus (PRSV) is one of the most devastating viruses of papaya that has significantly hampered papaya production across the globe. Although PRSV resistance is known in some of its wild relatives, such as *Vasconcellea cauliflora* and in some of the improved papaya genotypes, the molecular basis of this resistance mechanism has not been studied and understood. Plant microRNAs are an important class of small RNAs that regulate the gene expression in several plant species against the invading plant pathogens. These miRNAs are known to manifest the expression of genes involved in resistance against plant pathogens, through modulation of the plant's biochemistry and physiology. In this study we made an attempt to study the overall expression pattern of small RNAs and more specifically the miRNAs in different papaya genotypes from India, that exhibit varying levels of tolerance or resistance to PRSV. Our study found that the expression of some of the miRNAs was differentially regulated in these papaya genotypes and they had entirely different miRNA expression profile in healthy and PRSV infected symptomatic plants. This data may help in improvement of papaya cultivars for resistance against PRSV through new breeding initiatives or biotechnological approaches such as genome editing.

KEYWORDS

papaya, virus, PRSV, resistance, miRNA, small RNA, differential expression

Introduction

Papaya (*Carica papaya* L.), belonging to the family Caricaceae, is one of the economically most important nutraceutical fruit crops grown in tropical and subtropical regions of the world. Worldwide, the ring spot disease of papaya caused by the Papaya ringspot virus (PRSV) and transmitted by aphid vectors in a non-persistent manner, has

been the most devastating threat to papaya cultivation (Gonsalves et al., 2010; Mishra et al., 2016). PRSV, with a single-stranded positive sense RNA genome of 10 Kb length, encoding for a polyprotein that is eventually cleaved into 10 distinct proteins, belongs to the family *Potyviridae* and genus *Potyvirus* (Tripathi et al., 2008). The ringspot disease of papaya is identified by distorted leaves with mosaic pattern, wet oil streaks on petioles, stunted growth and ring spots on the fruits at later stages (Tripathi et al., 2008; Mishra et al., 2016).

Visible success has not been achieved in managing the PRSV under field conditions in many papaya growing regions of the world, mainly due to lack of agronomically accepted PRSV resistant papaya cultivars. Hence, developing PRSV resistant papaya varieties, either through conventional breeding or through biotechnological interventions, with acceptable agronomic traits and quality fruits is a priority (Mishra et al., 2016). Successful implementation of such virus resistant breeding programs, largely depends on the identification and selection of superior papaya genotypes that are resistant to PRSV (Prakash and Singh, 2013). Singh et al. (2006) conducted the study to identify superior papaya varieties that are tolerant to viral diseases under north Indian climatic conditions, by evaluating several accessions of papaya comprising of both Indian and exotic cultivars, along with promising selections. In an effort to develop PRSV resistant papaya, ICAR-Indian Agricultural Research Institute, Regional Station, Pune (IARI, RS, Pune, India) has successfully developed four PRSV tolerant dioecious lines named Pune Selection-1 (PS-1), PS-2, PS-3 and PS-5, by selecting and sib mating from segregating population of a land race of papaya called as Madhubala. All the four PS lines (-1, 2, 3 and 5) proved superior in respect of fruit yield and PRSV tolerance as compared to other commercial papaya varieties, suggesting that these lines can be used as a source of PRSV tolerance (Datar et al., 2013; Sharma and Tripathi, 2019; Sharma et al., 2019).

During the course of evolution, plants have developed diverse resistance mechanisms against the invading viruses such as RNA silencing, immune receptor signaling, protein degradation, hormone mediated defense, innate antiviral immunity, translation repression, small RNA-mediated antiviral defense, dominant viral resistance genes, resistance to virus movement, autophagy, and cross protection (Loebenstein, 2009; Muthamilarasan and Prasad, 2013; Park et al., 2013; Sett et al., 2022, Calil and Fontes, 2017, Patil et al., 2021). The RNA silencing is mediated by various types of small RNAs such as siRNAs, microRNAs, tasiRNAs etc (Weiberg and Jin, 2015). Unlike animals, plants lack defense cells and rely on the capacity of every cell to distinguish and defend against the invaders. Micro RNA is a class of unique small endogenous, non-coding RNA that is involved in regulation of gene expression by binding to the target mRNA leading to mRNA cleavage, translational repression, mRNA de-adenylation or transcriptional silencing thereby controlling the expression of the translated product (Jones-Rhoades et al., 2006; Winter et al., 2012). The control for gene expression is based on complementary pairing at specific positions for the target (Rhoades et al., 2002). These microRNAs are responsible for specific temporal and spatial control of their gene targets. Depending on the different environmental conditions, they act as an alternative strategy for editing the transcriptome during various environmental stresses. The role of miRNAs in plant development and stress response is well established and also has a crucial role in plant-virus interactions (Liu et al., 2017). Some of the plant miRNA families are highly conserved through thousands of millions of years, whereas some of them evolved and diversified to become specific to certain plant species and their genotypes (Axtell and Bartel, 2005; Axtell et al., 2007; Grimson et al., 2008). With the advent of high-throughput sequencing, it has been possible to identify not only the conserved miRNAs, but also plant species-specific miRNAs (Fahlgren et al., 2007; Moxon et al., 2008). Several studies have implicated the role of microRNAs that respond to diverse biotic stresses in various plant species and also diverse plant genotypes of same plant species. The miRNA accumulation at various levels in response to virus infection has been reported by various studies (Bazzini et al., 2007; Naqvi et al., 2010; Khraiwesh et al., 2012; Kumar, 2014). The plant miRNAs and their corresponding target genes have been identified to be responsive to infection by diverse viruses such as cucumber green mottle mosaic virus (CGMMV), cowpea severe mosaic virus (CPSMV), mungbean yellow mosaic India virus (MYMIV), sugarcane mosaic virus (SCMV) and soybean mosaic virus (SMV) in cucumber, cowpea, common bean, black gram, corn and soybean (Yin et al., 2013; Kundu et al., 2017; Liang et al., 2019; Patwa et al., 2019; Martins et al., 2020).

Several Potyviruses are also reported to result in differential expression of miRNAs in the host plants. Tobacco etch virus (TEV) and Potato virus Y (PVY) representative of *Potyviridae* family show higher expression of miR166, miR171, miR159, and miR167 (Bazzini et al., 2007). Further the differential expression of miR160, miR169, miR164 and miR156 was observed in infected *Nicotiana benthamiana* leaf samples (Bazzini et al., 2007). Another study on PVY, reported higher expression of miR398, miR171, miR168, and miR156, while double infection of PVX-PVY showed synergistic effect on phenotype and resulted in higher expression of miR156 and miR398, while miR171 was downregulated (Pacheco et al., 2012). In spite of accumulation of certain miRNAs, the targets of the corresponding miRNA were upregulated, and SPL6-IV mRNA was suggested to be inhibited in PVX-PVY infection (Pacheco et al., 2012). TEV infection alone increases the expression of miR159 and miR168 in *N. benthamiana* (Várallyay et al., 2010).

Recent advances in the field of miRNA-mediated gene silencing have been reported to be applied in several agricultural crop species to counter diverse virus infections (Tiwari et al., 2014; Khalid et al., 2017). In order to understand the mechanism of PRSV tolerance in Pune Selection (PS-3) as compared to susceptible papaya genotypes (PM), in this study, miRNA-based regulatory network is investigated using high-throughput deep sequencing technology. For the first time we report here differential expression of microRNAs in response to PRSV infection in both PRSV tolerant and susceptible genotypes of papaya (*C. papaya*.).

Materials and methods

Evaluation of selected papaya genotypes for PRSV tolerance through artificial inoculation

To screen for PRSV resistance in diverse genotypes of papaya, thirty plants of each of the papaya genotypes such as Pune Selections (PS-1, PS -2, PS-3, and PS-5) and other commercial papaya cultivars (PM, Co2, Pusa Delicious, Pusa Dwarf, Madhu Bindu and Red Lady) and the papaya wild relative *Vasconcellea cauliflora* were planted in pots for artificial inoculation. To reconfirm the PRSV tolerance in the above listed genotypes of papaya, all the test lines were sap inoculated at 4–6 leaf stage using 0.1 M phosphate buffer, pH 7.5 with previously maintained PRSV isolate in a susceptible papaya cv Red Lady in the insect proof glasshouse at IARI Regional Station, Pune (Gorane et al., 2019). Sap inoculated plants were kept in insect proof glasshouse for 4–6 weeks for recording the symptoms and further confirmation and quantification of PRSV.

Detection and quantification of PRSV by ELISA and RT-PCR

The leaf tissues from fully expanded young leaves of two papaya genotypes, PM (Susceptible) and PS 3 (tolerant) were collected from healthy and PRSV infected plants maintained at ICAR-IARI, Regional Station, Pune (Maharashtra, India). Leaves of the PRSV resistant wild relative of papaya *V. cauliflora*, were also collected for comparison. The presence and absence of PRSV infection in collected leaf samples was confirmed by ELISA and RT-PCR using PRSV specific antisera and primers, respectively. In order to confirm the virus infection, leaf samples from papaya plants were collected 4 weeks post inoculation and subjected to double antibody sandwich-enzyme linked immunosorbent assay (DAS-ELISA) using PRSV specific polyclonal antibodies (Agdia Inc., USA). Further reconfirmation of inoculated samples was done by reverse transcriptase polymerase chain reaction (RT-PCR) on selected samples of each papaya line. Total RNA was extracted from selected leaf samples by RNeasy plant mini kit (Qiagen, Germany) and cDNA synthesized using iScript™cDNA synthesis kit (Biorad, USA) following manufacturer's instructions. The PRSV CP gene was amplified by PCR using 2 µl cDNA template, 12.5 µl of PCR master mix (ThermoFisher Scientific, USA) and 10pM of each PRSV CP specific primers (Gorane et al., 2021). PCR program was setup as follows: 1 cycle of initial denaturation for 3 min at 94°C followed by 35 cycles of denaturation at 94°C for 30 sec, annealing at 60°C for 1 min and extension at 72°C for 1 min and 1 cycle of final extension at 72°C for 10 min. The PCR products were subjected to electrophoresis in a 1% (w/v) agarose gel, stained with GelRed nucleic acid stain (Biotium, USA) and visualized in Gel Documentation system (Syngene, UK).

RNA extraction, library construction and sequencing

Total RNA was extracted from the papaya leaves using Sigma Plant RNA isolation kit (Sigma-Aldrich, USA), according to the manufacturer's instructions. The small RNA library was constructed by the Small RNA Sample Pre-kit. The small RNA samples were ligated by using chimeric oligonucleotides 5'-GTTCAGAGTTCTACAGTCCGACGATC-3' and 3'-TCTGC ACACGAGAAGGCTAGA-5' with adapter sequences. The final cDNA library was ready for sequencing, after a round of adapter ligation, reverse transcription, PCR enrichment, purification and size selection. The qualified library was subsequently sent for Illumina HiSeq2500 sequencing to Nucleome Informatics Pvt. Ltd. (Hyderabad, India). The raw reads generated were processed by removing the adapter sequences, empty reads, no insert tags, oversized insertion, low quality reads (>50% of the bases with a quality score =5), poly A tags and small tags to obtain the clean reads (Supplementary Figure 1). The clean reads were further filtered based on read length of 18–24 nucleotides using in-house scripts.

Identification and differential expression analysis of miRNAs

Clean reads were first aligned with PRSV P isolate DEL (Accession No. EF017707, Parameswari et al., 2007) in which unaligned reads were considered for further analysis. Unaligned reads were mapped to the *C. papaya* hairpin loop of miRBase v21.0 (https://mirbase.org/) to filter known miRNAs. Known miRNA count files and unaligned reads from the above process were submitted to miARma-seq (miRNA-Seq and RNA-Seq Multiprocess Analysis) suite to predict novel miRNAs. This suite was used to identify differential expression analysis of known and novel miRNAs. The miARma-Seq (Andrés-León et al., 2016) uses edgeR software for differential expression analysis which includes: Read counts Normalisation, Model dependent p-value estimation and FDR value estimation based on multiple hypothesis testing. The significant differentially expressed known and novel miRNAs were filtered based on FDR < 0.05.

Target identification and functional annotation

The targets of some of the differentially expressed plant miRNAs were identified through the online tool psRNATarget (released 2017; https://www.zhaolab.org/psRNATarget/home) by using sequence of papaya (*C. papaya*) transcript, JGI genomic project, Phytozome, phytozome v8.0, internal number 113 (Dai et al., 2018). The psRNATarget server runs on a Linux cluster with a robust distributed computing back-end pipeline and is developed for high-throughput analysis of the NGS data. This software

encompasses latest findings of plant miRNA target recognition, that can distinguish the translational and post-transcriptional inhibition, thus with an ability to report the number of small RNA/target site pairs that can affect small RNA binding activity to target transcript. The psRNATarget evaluates complementarity between small RNA and target gene transcript using the scoring scheme originally applied by miRU (Zhang, 2005). The targets with single base or no mismatches in the seed region (position 2–8 bases) were considered for further analysis. The functional identification of targets was performed using functional annotation tools agriGo (Gene Ontology Analysis Toolkit and Database for Agricultural Community) (Tian et al., 2017).

Validation of small RNA data by RT-PCR for selected miRNAs

The small RNA data obtained from different papaya samples by next generation sequencing was validated using semi-quantitative RT-PCR (Figure 1). The total RNA was extracted from the papaya leaf samples using Spectrum Plant Total RNA Kit (Sigma-Aldrich, St. Louis, MO, USA) following the manufacturer's instructions. About 2 µg of total RNA was subjected to reverse transcription using the MultiScribe Reverse Transcriptase, RT Random Primers and other components from the High-Capacity cDNA Reverse Transcription kit (Applied Biosystems, Foster City, CA, USA). The RT-PCR reactions were carried out in a total reaction volume of 20 µL, with a cDNA synthesis cycle of 25 °C for 10 min, 37 °C for 120 min and 85 °C for 5 min, set up in a Bio-Rad S1000 Thermal cycler (Bio-Rad Laboratories, Hercules, CA, USA). A 1:10 dilution of the cDNA was used for PCR amplification of various miRNAs (miR160, miR164 and 5S rRNA), using specific primer pairs (Supplementary Table 1). The RT-PCR products were resolved in 2% TAE (Tris-acetate-EDTA) agarose gels stained with ethidium bromide and visualized under ultraviolet (UV)-light.

Results

Evaluation of papaya lines for tolerance to papaya ringspot virus

A total of 11 papaya genotypes (6 commercial cultivars, 4 Pune Selections (PS) and one wild relative of papaya *V. cauliflora*) were evaluated for resistance to PRSV, at IARI, Regional Station Pune (Maharashtra state, India). Overall, the PRSV-disease incidence recorded was highest (100%) in commercial cultivar PM and lowest (40%) in PS5, followed by PS3. The overall average PRSV disease incidence was higher (73%) in commercial cultivars as compared to PS lines where average disease incidence was 60%. The presence of PRSV infection was confirmed in all the test genotypes by both ELISA and RT PCR. However, the virus titer varied in different papaya lines. The highest ELISA OD value of 1.55 was recorded in Honey Dew and the lowest value of 0.42 in PS3 papaya line. These OD values varied from 0.8 to 1.55 in commercial papaya test lines. Overall, the average ELISA value was recorded higher (1.29) in commercial papaya cultivar as compared to PS lines (0.83). Moreover, the field experiments over the years for evaluation of PRSV resistance showed similar pattern of disease intensity as reported by Sharma and Tripathi (2019). The above finding along with previous reports of field experiments showed that the PS lines have a good level of tolerance against PRSV infection as compared to the other commercial papaya cultivars (Table 1; Figure 2).

Analysis of small RNA sequencing data

To study the impact of PRSV infection with reference to different miRNAs and transcriptomic changes at genome-scale, two different papaya varieties PM (PRSV susceptible) and PS3 (PRSV tolerant) were used along with *V. cauliflora* (VC; PRSV resistant). For further characterization, Illumina HiSeq2500

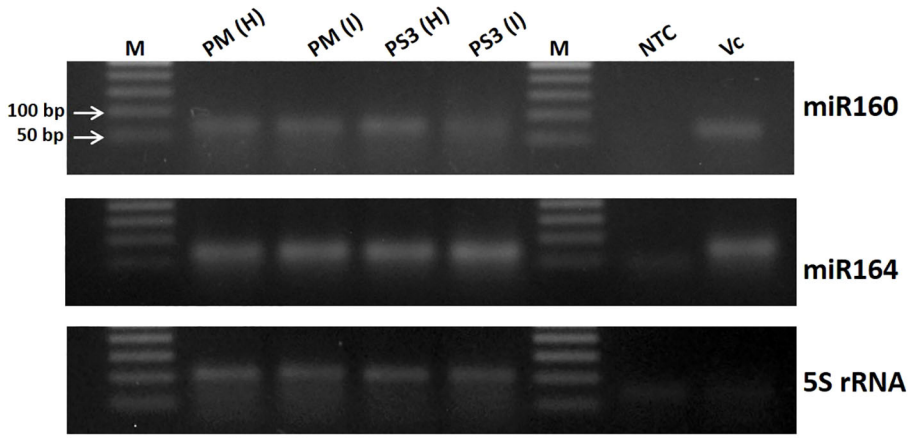


FIGURE 1
Validation of small RNA data obtained from different papaya genotype leaf samples by next generation sequencing using semi-quantitative RT-PCR. M, DNA ladder; PM-H, Pusa Majesty-Healthy; PM_I, PM-Infected; PS3_H, Pune Selection 3 – Healthy; PS3_I, Pune Selection 3 – Infected; NTC, Negative Control and VC, *Vasconcellea cauliflora*.

TABLE 1 Field evaluation of different papaya genotypes for tolerance to Papaya ringspot virus (PRSV) infection and estimation of PRSV titer by ELISA and RT-PCR.

Papaya genotype/cultivar	PRSV infectivity in glasshouse (%)	PRSV titer estimation by ELISA (OD value)	PRSV confirmation by RT-PCR	Type of PRSV Symptoms observed
Red Lady	80	++ (1.51)	+	Mild Mosaic
Honey Dew	60	++(1.55)	+	Mosaic
Pusa Dwarf	80	+ +(1.51)	+	Mosaic
Pusa Delicious	80	+ (1.06)	+	Mosaic
Co-2	40	+ (0.8)	+	Mild mosaic
PM	100	+ (1.29)	+	Mosaic
V.C.	0	–	–	No Symptoms
PS-3	60	+ (0.42)	+	Mild mosaic
PS-1	60	+(1.14)	+	Mild mosaic
PS-2	80	+ (0.96)	+	Mild mosaic
PS-5	40	+ (0.80)	+	Mild mosaic

Pusa Majesty (PM), Pune Selections (PS)-1, 2, 3 & 5, and *Vasconcellea cauliflora* (VC). Higher virus titer (++), Moderate virus titer (+), Virus not detected (-).

sequencing platform was used to sequence the healthy and infected PM and PS3, together with VC, which generated a total of 237.8 million reads. Comparable numbers of reads were obtained in healthy samples of PRSV susceptible (PM) and a tolerant (PS3) variety i.e. 52 million reads while VC and infected samples ranged from 42–45 million reads (Table 2).

A high percentage (99%) of clean reads were obtained after further processing, filtering and mapping of raw reads with the reference genome. These clean sequences were further classified as unaligned reads and virus aligned reads (Table 2). The unaligned sequences were varying in different samples, 71% clean reads were unaligned in PM(I), 50% in PS3(I), 40% in PS3(H), 37.7% in VC, and 22.6% in PM(H) (Table 2). A maximum of 4.3 million reads of novel miRNAs was obtained after mapping unique and unaligned reads for PM(I), whereas PS3(I) had only 0.711 million reads of novel RNAs (Table 2).

The size distribution of the sequenced small RNA ranged from 18 to 24 nucleotides (nt) (Figure 3). The abundance of small RNA reads is not identical in all the samples, 21nt sequence was predominant in PRSV infected samples of PM and PS3, while 23nt was dominant in PM-H, 24nt were dominant in PS3-H and VC. Whereas the smaller sized small RNA species i.e., 18–20nt were in reduced amounts in all the plant samples irrespective of PRSV infection status (Figure 3).

Identification and quantification of known and novel miRNAs

In order to filter known miRNAs, the unique sequences were further mapped to the *C. papaya* hairpin loop of miRBase v21 (https://mirbase.org/) and were quantified using in-house scripts. According to the alignment results, the maximum number of reads were noted in the PM (I) sample i.e. 1,145,855 while the healthy samples PM(H) showed 160,874 reads (Table 2; Figure 4A). The wild relative of papaya *V. cauliflora* showed the lowest reads of 114,025.

It was interesting to note that the PRSV tolerant papaya variety PS3 showed higher reads in healthy samples as compared to the PRSV infected samples (Table 2). A total of 79 known miRNAs were identified among different samples, 28 known miRNAs show more than 1000 reads (Figure 4A). Maximum number of reads were recorded for the two microRNAs, miR159a and miR398 (Figure 4A; Table 3). A comparison between PRSV resistant and susceptible papaya varieties suggests that the 18 miRNAs were differentially expressed in infected PS3 and PM samples, while 9 miRNAs were differentially expressed in healthy samples and 6 miRNAs were present in both healthy as well as the infected samples (Figure 4C). Further, the infected and healthy samples were compared with each other, wherein 15 known miRNAs were identified in PM variety that were not present in PS3, while 4 miRNAs were differentially expressed in PS3 alone (Figure 4B). On the other hand, 29 known miRNAs were differentially expressed in both the varieties (Figure 4C). The known miRNA count files and unaligned reads from the above process were submitted to miARma-seq suite to predict novel miRNAs (mirDeep of MiARma-Seq suite). A total of 291 novel miRNAs were identified in different papaya genotypes (Table 4).

Differential expression of miRNAs associated with PRSV infection

A comparative analysis of change in the levels of expression of miRNAs in the infected samples vs healthy samples indicated that the higher number of miRNAs were upregulated in infected samples. A similar pattern was noticed in both known as well as novel miRNAs (Table 4).

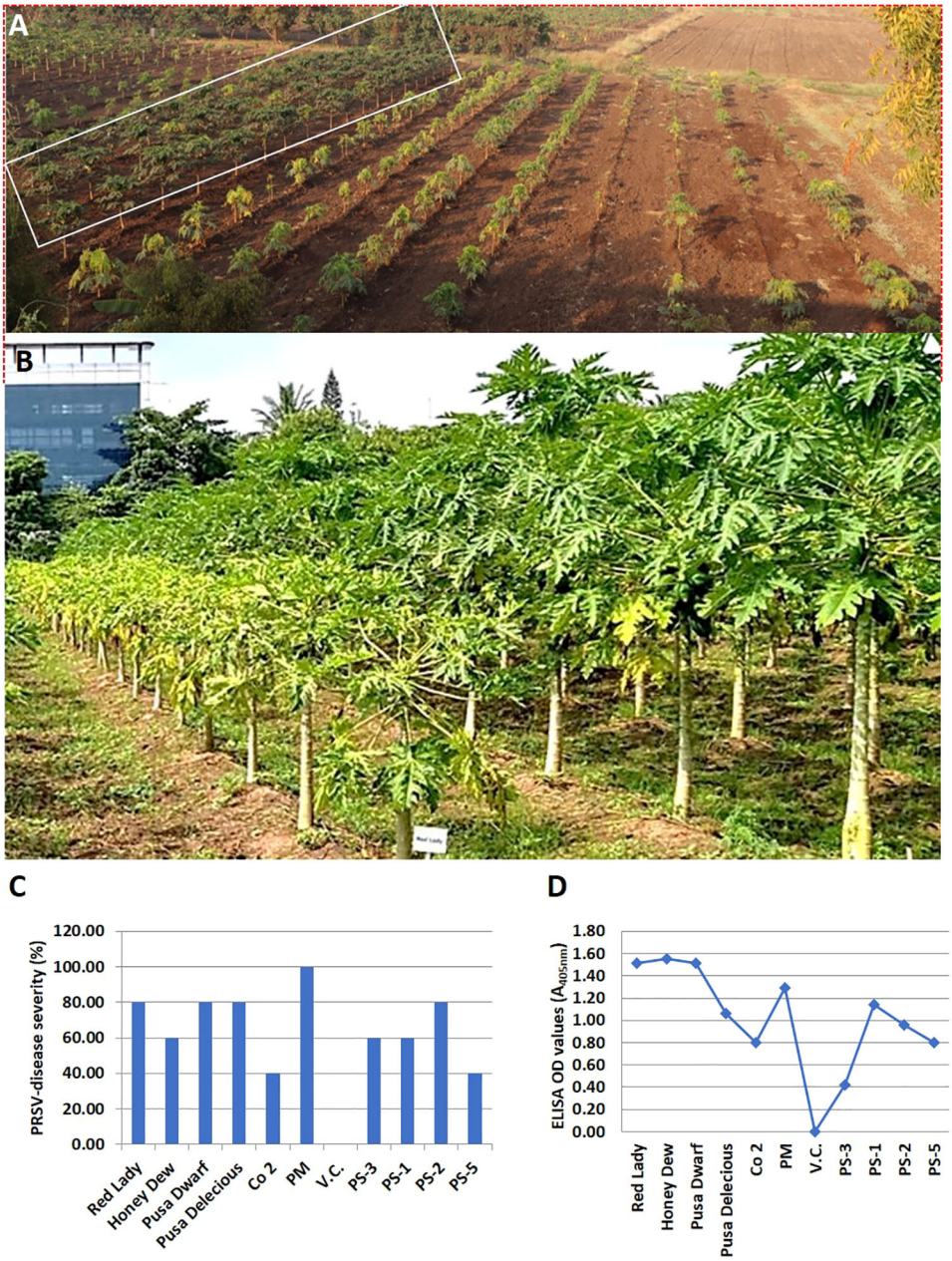


FIGURE 2
Performance of various papaya genotypes for PRSV resistance at the IARI experimental Station, Pune, Maharashtra state, India. **(A)** The papaya plant rows that are marked in a rectangular box are the PRSV tolerant Pune selection (PS) papaya lines, while the papaya plant rows on its right side are the PRSV susceptible commercial papaya varieties. **(B)** Performance of PRSV tolerant Pune selection line (right row) vs susceptible commercial variety (left row) at fruiting stage. **(C)** Levels of PRSV infectivity, in percentage, as exhibited by different papaya genotypes in glasshouse by artificial PRSV inoculation, **(D)** Graphical representation of PRSV load in artificially inoculated papaya genotypes under glasshouse condition, as determined by ELISA (OD value).

Known miRNAs

A total of 44 miRNAs were compared of which 24 were upregulated and 20 were downregulated in the PRSV infected susceptible papaya variety PM. The heat map of the data clearly suggests that the conserved miRNAs belonging to family miR395, miR171, miR160, miR156, miR172, miR169, miR164, miR396, miR398, miR408, miR167 and miR477 show differential expression in infected and healthy samples (Figure 5). Amongst the conserved miRNAs, miR396, miR160f, miR160c, miR390b, miR390a, miR393, miR398, miR164b, miR156a, miR156b, miR156c, miR171d, miR395b, miR395a, miR395d, miR395e, miR162a, miR319, miR535 were upregulated, whereas miR477, miR166c, miR166d, miR166b, miR408, miR159a, miR167d, miR167c, miR167b, miR167a, miR160e, miR171b, miR159b and miR394b were downregulated in PRSV infected PM samples as compared to PM-H. Some of the miRNA specifically present in

TABLE 2 Summary of type and number of small RNA reads obtained by Illumina HiSeq 2500 sequencing in healthy and infected Papaya varieties PM (PRSV susceptible), PS3 (PRSV tolerant) and VC (Wild relative of papaya).

	Healthy Papaya Plants			PRSV infected Plants	
RNA class	PM (H)	PS3 (H)	VC	PM (I)	PS3 (I)
Raw reads	52,989,070	52,663,396	42,988,993	43,368,829	45,873,913
Total Clean reads	52,841,249	52,543,607	42,876,839	43,262,152	45,749,345
Clean reads (18–24 nt)	12,156,513	21,983,529	16,305,777	34,217,191	24,350,232
Unaligned reads	11,994,524	21,021,736	16,164,592	31,146,449	22,893,062
Novel miRNA reads	139,567	419,784	114,025	4,371,533	711,008
Known miRNA reads	160,874	943,881	139,427	1,145,855	688,181
Virus alignment	1,115	17,912	1,758	1,924,887	768,989

Pusa Majesty (PM), Pune Selections-3 (PS- 3), and *Vasconcellea cauliflora* (VC); H, healthy; I, PRSV infected.

papaya such as miR8137, miR8140, miR8154, miR8135 are upregulated while miR8148, miR8152, miR8153, miR8149 and miR8144 were downregulated in infected samples as compared to healthy samples of PM (Figure 5).

The PRSV tolerant variety (PS3) shows contrasting expression as compared to susceptible papaya variety (PM) upon PRSV infection. The miRNAs miR408, miR166b, miR166c, miR166d, miR159a, miR477, miR167d, miR167c, miR8148, miR8152, miR8149, miR8153 are upregulated in PS3-I, whereas they get downregulated in PM-I when compared to their corresponding healthy samples (Table 3). Similarly, contrasting expression pattern was noted for the miRNAs: miR396, miR160f, miR160c, miR390b, miR390a, miR393, miR164b, miR156a, miR156b, miR156c, miR171d, miR395b, miR535, miR8137, miR8140, which gets downregulated in infected samples of PS3, while upregulated in PM-I. The miR398 is upregulated in infected samples of both the papaya varieties (Table 3).

Apart from contrasting results the miRNA identified only in PS3 variety includes miR160d which is upregulated, and miR164a, c, d that is downregulated in PRSV infected papaya samples (Table 3). A comparison between healthy samples of the papaya varieties PM and PS3 suggests differential expression of 15 miRNAs, among them the 2 miRNAs, miR162a and miR8155 were upregulated while 13 were downregulated (Table 3). The PRSV infected sample PS3 when compared with PM-I showed higher expression of 10 miRNAs while lower expression of 14 miRNAs. The miR160 a,b and miR171c were present only in the infected samples of PS3 and not in PM-I (Table 3).

Another comparison of miRNA expression was made between the PM, PS3 with the wild relative of papaya VC (PRSV resistant). Interestingly both the papaya varieties PM and PS3 showed similar miRNA expression pattern (Table 3). The expression of miR156 a,b, c, miR160c,d, miR160f, miR164a-e, miR167d, miR390a,b, miR393, miR396, miR398, miR408, miR477, miR5211, miR8135–37,

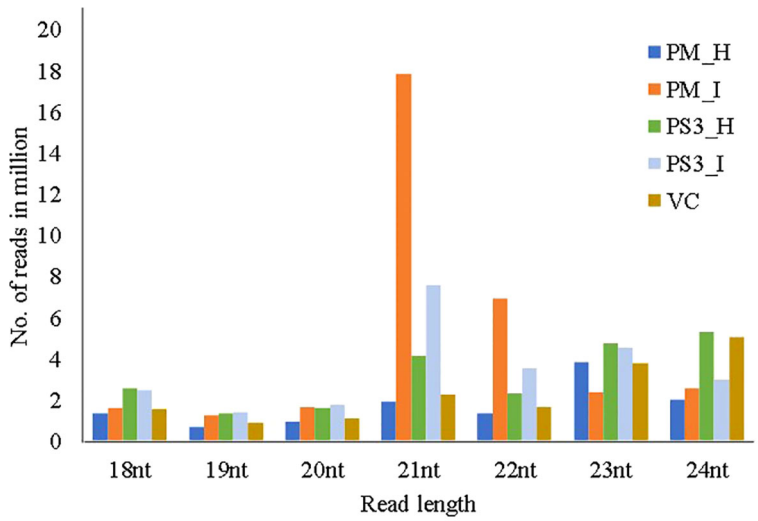


FIGURE 3
Size distribution of small RNAs, with read lengths in the range of 18–24nts, retrieved through Illumina next generation sequencing from different papaya genotypes that show different levels of tolerance to PRSV. PM-H, Pusa Majesty-Healthy; PM_I, PM-Infected; PS3_H, Pune Selection 3 – Healthy; PS3_I, Pune Selection 3 – Infected; and VC, *Vasconcellea cauliflora*.

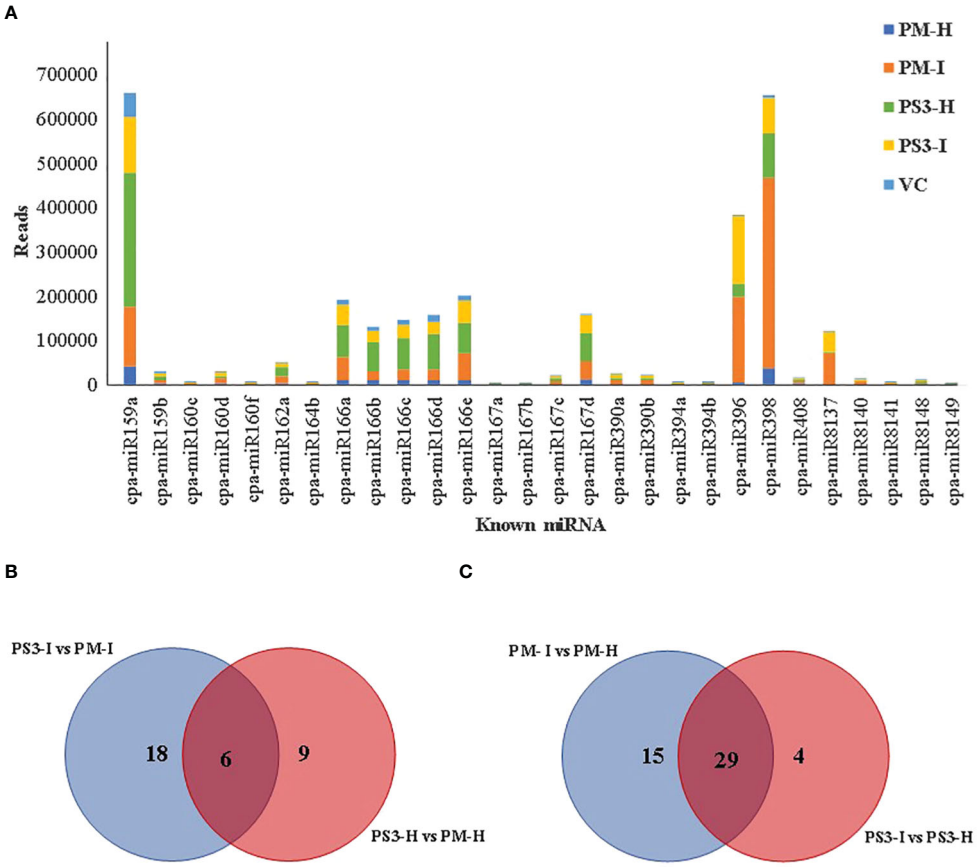


FIGURE 4
Differentially expressed miRNAs between PRSV infected and non-infected papaya genotypes: **(A)**. The known miRNAs showing more than 1000 reads in any one of the samples, **(B)** and **(C)** Venn diagrams showing known differentially expressed miRNAs in Healthy and infected, susceptible and resistant papaya cultivars. PM-H, Pusa Majesty-Healthy; PM_I, Pusa Majesty-Infected; PS3_H, Pune Selection-3 – Healthy; PS3_I, Pune Selection-3 – Infected; and VC, *Vasconcellea cauliflora*.

TABLE 3 Differential expression of known miRNAs in healthy and PRSV infected Papaya plant leaf samples and their comparative expression is denoted as ↑ (boxed in green color) as upregulation, ↓ (boxed in red color) as downregulation.

	PM- I vs PM-H	PS3-I vs PS3-H	PS3-H vs PM-H	PS3-I vs PM-I	PM-I vs VC	PS3-I vs VC	VC vs PM-H	VC vs PS3-H
cpa-miR156a	↑	↑	–	↓	↑	↑	↓	↓
cpa-miR156b	↑	↑	–	↓	↑	↑	↓	↓
cpa-miR156c	↑	↑	–		↑	↑	↓	↓
cpa-miR156e					↓	↓	↑	↑
cpa-miR156f					↓	↓	↑	↑
cpa-miR159a	↓	↓	–		↓	↓	↑	
cpa-miR159b	↓	–	–		↓	↓	↑	↑
cpa-miR160a	–	↑	↓	↑	↓	↓	↑	↑
cpa-miR160b	–	↑	↓	↑	↓	↓	↑	↑
cpa-miR160c	↑	↑	↓		↑		↑	↑
cpa-miR160d	–	↑	↓		↑	↑	↓	↓
cpa-miR160e	↓	↑	↓	↑	↓	↓	↑	

(Continued)

TABLE 3 Continued

	PM- I vs PM-H	PS3-I vs PS3-H	PS3-H vs PM-H	PS3-I vs PM-I	PM-I vs VC	PS3-I vs VC	VC vs PM-H	VC vs PS3-H
cpa-miR160f	↑	↑	↓		↑	↑	↑	↑
cpa-miR162a	↑	–	↑				↑	
cpa-miR164a	–	↑	↓		↑	↑		
cpa-miR164b	↑	↑	–		↑	↑	↓	
cpa-miR164c	–	↑	↓		↑	↑	↓	
cpa-miR164d	–	↑	–			↑		
cpa-miR164e		↑			↑	↑		
cpa-miR166b	↓	↓	–		↓	↓		
cpa-miR166c	↓	↓	–	↑	↓	↓		
cpa-miR166d	↓	↓	–	↑	↓	↓	↑	↑
cpa-miR167a	↓	–	–					
cpa-miR167b	↓	–	–					
cpa-miR167c	↓	↓	–		↓			
cpa-miR167d	↓	–	–		↑	↑	↓	↓
cpa-miR169					↓	↓	↑	↑
cpa-miR171a					↓	↓	↑	↑
cpa-miR171b	↓	–	–	↑	↓	↓	↑	↑
cpa-miR171c				↑	↓	↓	↑	↑
cpa-miR171d	↑	↑	–		↓	↓	↑	↑
cpa-miR172a					↓			
cpa-miR172b					↓	↓	↑	↑
cpa-miR319	↑	–	–	↓	↓	↓	↑	↑
cpa-miR390a	↑	↑	–			↑		
cpa-miR390b	↑	↑	–		↑	↑	↑	↑
cpa-miR393	↑	↑	–		↑	↑		
cpa-miR394a	↓	↓	–		↓	↓		
cpa-miR394b	↓	–	–					↑
cpa-miR395a	↑	↑	–		↓	↓	↑	↑
cpa-miR395b	↑	↑	–		↓		↑	
cpa-miR395c		↑			↓	↓	↑	↑
cpa-miR395d	↑	↑	–	↓	↓	↓	↑	↑
cpa-miR395e	↑	↑	–	↓	↓	↓	↑	↑
cpa-miR396	↑	↑	–		↑	↑	↓	↓
cpa-miR398	↑	–	↓	↓	↑	↑	↓	↓
cpa-miR408	↓	↓	↓	↓	↑	↑	↓	↓
cpa-miR477	↓	↑	↓	↑	↑	↑	↓	↓
cpa-miR5211	↑	–	–		↑			
cpa-miR535	↑	↑	–					

(Continued)

TABLE 3 Continued

	PM- I vs PM-H	PS3-I vs PS3-H	PS3-H vs PM-H	PS3-I vs PM-I	PM-I vs VC	PS3-I vs VC	VC vs PM-H	VC vs PS3-H
cpa-miR8134	–	↑						
cpa-miR8135	↑	↑	–	↓	↑	↑	↓	↓
cpa-miR8136					↑	↑	↓	↓
cpa-miR8137	↑	↑	↓		↑		↓	↓
cpa-miR8139a				↓	↑	↑	↓	↓
cpa-miR8139b		↑				↑	↓	
cpa-miR8139c					↑			
cpa-miR8139d					↑	↑	↓	↓
cpa-miR8139e				↓	↑	↑	↓	↓
cpa-miR8140	↑	↑	–		↓	↓	↑	↑
cpa-miR8141					↑	↑	↓	↓
cpa-miR8142					↑	↑	↓	↓
cpa-miR8143		↑				↑		
cpa-miR8144	↓	↓	–	↑	↑	↑	↓	↓
cpa-miR8145					↑	↑	↓	↓
cpa-miR8146				↓	↑	↑	↓	↓
cpa-miR8148	↓	↓	–	↑	↑	↑	↓	↓
cpa-miR8149	↓	–	–		↑	↑	↓	↓
cpa-miR8150		↓			↑	↑	↓	↓
cpa-miR8151					↑	↑	↓	↓
cpa-miR8152	↓	–	↓		↑	↑	↓	↓
cpa-miR8153	↓	–			↑	↑	↓	↓
cpa-miR8154	↑	↑		↓	↑	↑		
cpa-miR8155	–	↓	↑		↓	↓	↑	

Pusa Majesty (PM), Pune Selection-3 (PS3), and Vasconcellea cauliflora (VC).

miR8139a-e, miR8141–46, miR8148–54, were downregulated in healthy samples of both the papaya varieties. Some of the miRNAs which were downregulated in infected samples when compared to VC include miR156e-f, miR159a-b, miR160a,b, miR160e, miR166b-d, miR167c, miR169, miR171a-d, miR172a-b, miR319, miR394a, miR395a-e, miR8140, and miR8155 (Table 3).

Novel miRNAs

A total of 291 novel miRNAs were identified in the infected and healthy papaya leaf samples of contrasting genotypes and the wild relative. The sequencing reads suggest differential expression of 52 and 43 novel miRNAs in PM and PS3 respectively. Amongst these 29 miRNAs were upregulated and 8 miRNAs were downregulated in infected samples of both the papaya varieties, while 5 miRNAs showed differential expression. Two miRNAs unique for the papaya variety PM-I were downregulated, while one unique miRNA was upregulated in PS3-I samples. The PRSV resistant *V. cauliflora* samples were also sequenced and a comparison was made with the healthy samples of both the papaya varieties (Table 3). Twenty-one miRNAs were differentially expressed, 10 miRNAs showed similar expression, 2 miRNAs were upregulated, and 8 miRNAs were downregulated in VC when compared with PM and PS3 varieties (Table 3).

Target analysis of miRNAs

The miRNAs are negative regulators of gene expression, therefore, to investigate the role of miRNA, it is important to predict the target transcript sequences. The conserved miRNAs are associated with conserved targets that are common in various plant species such as Arabidopsis, rice, and poplar (Liang et al., 2013). The subfamilies of miRNAs differing at 1–2 nucleotide positions target same transcript, miR160a-f targets auxin response factors (ARF10, 16, 17), miR156a-c targets Squamosa promoter-binding-like protein 7,2, and the miR166b-d targets homeobox leucine zipper protein. Whereas the miR164a-c targets NAC domain containing protein, and the miR171b-d targets

Scarecrow-like protein, histone lysine N-methyltransferase ATX2-like protein. The miRNAs are downregulated in VC as compared with papaya, that targets transcription factors of family SPL, ARF, NAC domain, growth factors (Supplementary Table 2). These transcription factors are involved in leaf development and lateral root development. The CpARF10, CpARF16, and CpARF17 showed fruit-specific expression, which indicated that they might play an important role in fruit ripening (Liu et al., 2015a).

Discussion

MicroRNAs (miRNAs), are long noncoding RNAs, that regulate gene expression by binding to the target mRNAs, resulting in mRNA cleavage or inhibition of protein translation (Ambros, 2001; Carrington and Ambros, 2003; Bartel, 2004). The plant miRNAs play a crucial role in diverse biological phenomenon, such as development of leaf (Rhoades et al., 2002; Chen, 2005; Guo et al., 2005; Lauter et al., 2005), nutrient homeostasis (Chiou, 2007) and various stress responses (Chapman et al., 2004; Jones-Rhoades and Bartel, 2004). Understanding the differential expression pattern of miRNAs and their role during the virus infection in papaya will provide the clues to design advanced strategies to control the PRSV infection in papaya and other high value horticultural crops.

Therefore, in this study, we present a detailed analysis of the miRNA expression pattern in papaya plants that are susceptible (PM) and tolerant (PS3) to PRSV, with or without PRSV infection/symptoms, along with the wild relative of papaya *V. cauliflora*, to understand the miRNA-mediated PRSV resistance. We performed deep sequencing by employing Illumina HiSeq 2500 sequencing and compared the miRNA expression between the resistant and susceptible papaya genotypes, following PRSV infection. Here we also report and reconfirm similar findings of PRSV tolerance in papaya lines that was previously demonstrated by Sharma and Tripathi (2019), wherein the Pune Selection (PS) lines showed significant level of PRSV tolerance compared to the commercially cultivated papaya cultivar (PM: Pusa Majesty) under the field trials conducted for more than four years, at Pune in Maharashtra state (India). The present study is based on the past reports of Chavan et al. (2010), wherein eight commercial papaya cultivars were screened for PRSV resistance and a lowest PRSV incidence of 13.2%, was recorded for the papaya variety 'Madhubala', that subsequently served as parental material for the PRSV tolerant Pune Selections (PS). Both the present and the past studies reported a better tolerance to PRSV in PS lines when evaluated in different climatic conditions, at different time points. Thus, these PS lines can be a potential source of PRSV tolerance and can be incorporated in the future PRSV resistance breeding programmes. Further, *V. cauliflora*, that are wild relatives of cultivated papaya, were found to be immune to PRSV infection, which could be because of its natural immunity or non-host nature. As expected, PRSV was not detected in virus inoculated *V. cauliflora* plants, either by ELISA or by RT-PCR, which is in conformity with the previous reports (Sharma and Tripathi, 2019; Sharma et al., 2019). Based on the PRSV infection data from present and previously reported studies, PS3 can be a good candidate to further investigate the molecular basis of virus tolerance. Therefore, the two contrasting genotypes with reference to PRSV tolerance, PS3 and PM were selected to study the differential expression of known and novel miRNAs in response to PRSV infection, by employing next generation sequencing technologies. Some of the miRNA targets found in papaya are involved in the ubiquitin-dependent protein catabolic process and the increasing evidence indicates that plants utilize this process during their immune response to pathogen invasion. However, this ubiquitin pathway can also be used by the viruses to enhance the infection process, by enhancing their own replication (Marino et al., 2012). Infection of *N. tabacum* by the RNA viruses Tobacco etch virus (TEV) and Potato virus Y(PVY) representative of the *Potyviridae* family show higher expression of miR166, miR171, miR159, and miR167 post infection; while differential expression of miR160, miR169, miR164 and miR156 in TEV and PVY infected plants (Bazzini et al., 2007). Similarly, studies on another member of *Potyviridae* revealed that miR160, miR393, and miR1510 were involved in resistance to SMV infection in soybean plants (Yin et al., 2013). The small RNA (sRNA)-sequencing, degradome-sequencing, as well as a genome-wide transcriptome analysis in SMV infected soybean revealed that increase in the expression level of miR168 leads to a serious

TABLE 4 Differential expression of known and novel miRNAs in healthy and infected papaya varieties: Pusa Majesty (PM), Pune Selection-3 (PS3), and *Vasconcellea cauliflora* (VC).

	Known miRNAs			Novel miRNAs		
Samples	Up	Down	Total	Up	Down	Total
PM-I vs PM-H	24	20	44	37	15	52
PM-I vs PS3-H	36	17	53	35	14	49
PM-I vs VC	38	27	65	40	8	48
PS3-I vs PM-H	18	15	33	40	6	46
PS3-I vs PS3-H	33	11	44	34	9	43
PS3-I vs VC	39	25	64	40	7	47
VC vs PM-H	26	29	55	7	14	21
VC vs PS3-H	23	27	50	4	17	21

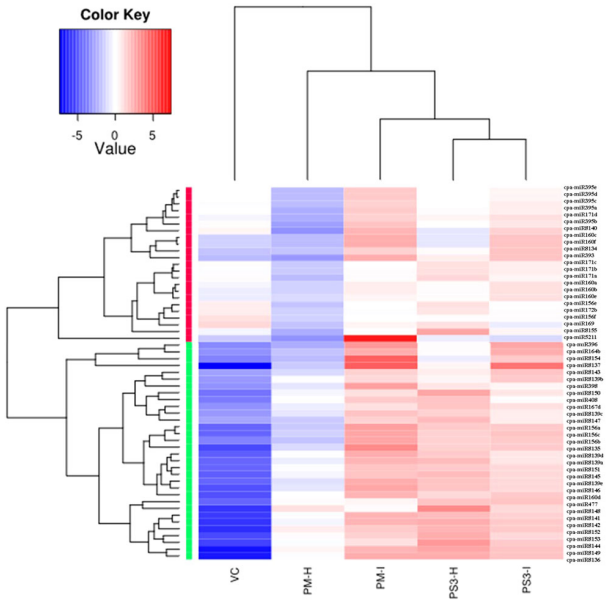


FIGURE 5
Heat map obtained from sequencing data showing differentially expressed known miRNAs in PRSV infected leaves (PM-I and PS3-I) to healthy samples (PM-H, PS3-H) and VC. cpa-miR stands for microRNA of *Carica papaya* plant. The color scale shown at the top illustrates relative expression of a miRNA across all samples: red represents expression above mean, blue represents expression lower than mean. PM-H, Pusa Majesty-Healthy; PM_I, PM-Infected; PS3_H, Pune Selection 3 – Healthy; PS3_I, Pune Selection 3 – Infected; and VC, *Vasconcellea cauliflora*.

inhibition of the target AGO1 mRNA, that encodes for a member of the argonaute family of proteins, which associate with the small RNAs and have important roles in RNA interference (Chen et al., 2015, 2016). In another study, microarray analysis indicated that the up-regulated miRNAs, namely, miR168a, miR403a, miR162b and miR1515a regulated the expression of AGO1, AGO2, DCL1 and DCL2, that encode for the components of silencing complex (Bao et al., 2018). Studies by Abreu et al. (2014) on response of papaya microRNAs to infection by papaya meleira virus (PMeV) resulted in identification of 462 microRNAs, representing 72 microRNA families. The expression of 11 microRNAs, with potential targets in 20S and 26S proteasomal degradation pathway was studied in response to PMeV infection and its titer.

In the present study, the abundant reads of small RNA (23nt and 24nt) were predominant in PM-H, PS3-H and *V. cauliflora*. Whereas, the 21nt small RNA reads were predominant in PRSV infected leaf samples of PM and PS3. These results are similar to earlier reports which have shown that 24nt small RNAs are more abundant in several other plant-pathogen interactions such as tomato-late blight, tomato-cucumber mosaic virus (CMV) and wheat-powdery mildew pathogen (Xin et al., 2010; Feng et al., 2014; Luan et al., 2015). The distribution of different sized distinct small RNAs may reflect their compositions (Ding and Lu, 2011). In this study, we found that both 23nt and 24nt small RNAs were more abundant in the healthy papaya plants than in the PRSV infected papaya. Whereas the 21nt and 22nt sized miRNAs were more abundant in both the papaya varieties, PM and PS3 infected with PRSV. This finding is in line with reports by Du et al. (2019) where they found larger number of 24nt sized small RNAs in the non-infected library than in the virus infected. The expression profile of miRNAs was altered on co-infection of tobacco curly shoot virus along with its betasatellite in *N. benthamiana* (Du et al., 2019). A number of siRNAs and their targets have been shown to determine leaf development, such as miR156-Squamosa Promoter Binding Protein-Like (SPL) (Xu et al., 2016), miR160-Auxin Response Factor (ARF) (Ben-Gera et al., 2016), miR165/166-Class III Homeodomain-Leucine Zipper (HD-ZIPIII) (Jia et al., 2015), miR319-Teosinte Branched/Cycloidea/Proliferating Cell Factor (TCP) (Bresso et al., 2018), miR390-Trans-Acting Small Interfering RNA3 (TAS3) (Husbands et al., 2016) and miR396-Growth Regulating Factor (GRF) (Omidbakhshfard et al., 2015). Additionally, multiple crucial components that are required for miRNA and ta-siRNA biogenesis, such as AGO1 (Bohmert et al., 1998), AGO7 (Garcia et al., 2006), SERRATE (SE) (Prigge and Wagner, 2001), and Hyponastic Leaves 1 (HYL1) (Liu et al., 2011) are suggested to influence the leaf development. The miRNA expression levels are associated with the different types of leaf curvatures (Williams et al., 2005). The qPCR of miRNAs revealed that a higher level of miR166 was associated with the downward curvature of leaf, while higher miR319a expression was correlated with wavy margins of the leaves (Williams et al., 2005).

In our study, a total of 44 known miRNAs and 291 potentially novel miRNAs were found to be differentially regulated during PRSV infection in papaya. Out of the 44 known miRNAs 20 miRNAs were downregulated and 24 were upregulated. Several studies report that the miRNAs also regulate host defenses against pathogens, including viruses, by suppressing pathogen multiplication at the post-transcriptional level (Al Abdallat et al., 2014; Feng et al., 2014; Du et al., 2019). Several miRNAs (e.g., miR168, miR169, and miR482) have been reported to target transcription factors controlling host resistance to virus infection (Várallyay et al., 2010; Zhu et al., 2013; Li et al., 2017). In *N. benthamiana*, virus infection may regulate the expression of

miR168 to alleviate the anti-viral function of AGO1 protein (Várallyay et al., 2010; Du et al., 2019). Our study also showed the differential expression of 15 miRNAs which might play a key role in resisting the PRSV infection, by manipulating some of the genes involved in metabolic or gene silencing pathway. Past studies had shown that the rice miR164 played an important role in resistance to southern rice black-streaked dwarf virus infection, as well as resistance to drought stresses by differentially regulating its target genes in rice crop (Feng et al., 2014; Xu et al., 2014). Du et al. (2019) had also shown the expression of miR164a and miR482 in the two libraries derived from *N. benthamiana* coinfected with TbCSV and TbCSB, suggesting that these two miRNAs may have an important role in virus resistance. Further, studies by Xiao et al. (2014) indicated for a role of miRNA-mediated production of phasiRNAs in interaction between begomovirus and the model host plant *N. benthamiana*. Various studies have demonstrated a role of miRNA-mediated phasiRNA pathway in diverse biological processes, including pathogen resistance (Kurihara et al., 2006; Kang et al., 2012; Peng et al., 2014; Cui et al., 2017; Zheng et al., 2020). The miR482/2118-mediated cleavage of NBS-LRR transcripts involved in disease resistance play an important role in non-race specific disease resistance and also triggers production of phased small RNAs that regulate the expression of their target genes (Kurihara et al., 2006; Kang et al., 2012). On infection by the virus, the expression level of miR482 was down-regulated and as a result the NBS-LRR transcripts were up-regulated (Kang et al., 2012; Cui et al., 2017). Analyzed 18 miRNA libraries prepared from SMV resistant and susceptible near-isogenic lines (NILs) of soybean, at three different time intervals of post virus infection. Their study found that a large number of miRNAs were differentially expressed in the two soybean NILs, that targeted a series of NBS-LRR resistance (R) genes. In a similar study by Soltani et al. (2021), the effect of CMV infection on different quinoa varieties, resulted in identification of differentially expressed miRNAs and the corresponding genes that modulated the variety-specific biological pathways, such as plant-pathogen interaction (PPI), DNA replication, repair and recombination, and hormone signaling. These biological pathways ultimately result in modulation of defense hormones such as SA, JA, and ET (Soltani et al., 2020). Some of the biological phenomenon, such as DNA replication and repair in the host genome are critical for antiviral mechanism (Soltani et al., 2020). Upregulation of DNA recombination resulted in persistent induction of LRR gene expression in Tobacco mosaic virus (TMV) infected tobacco plants ultimately leading to high level of resistance against TMV (Kovalchuk et al., 2003; Boyko et al., 2007).

Several studies have shown that the plant miRNAs target and negatively regulate the plant R genes by prompting the production of phased, trans-acting siRNAs (tasiRNAs) against these R genes, and this miRNA-mediated gene regulation is suppressed on bacterial or viral infection (Zhai et al., 2011; Li et al., 2012). In tobacco, the R gene "N" against TMV was found to be regulated by miR482 (Whitham et al., 1994; Li et al., 2012). In brief, the silencing of NBS-LRR genes by miR482, and their activation after miR482 down-regulation upon bacterial or viral infection, have been widely studied in different host plants (Li et al., 2012; Shivaprasad et al., 2012; Zhu et al., 2013). Similarly, Li et al. (2012) demonstrated that the expression of miR6019 and miR6020 in tobacco plants result in specific cleavage of transcripts of the N gene and its homologs by binding to the complementary sequence of the conserved Toll and Interleukin-1 receptors (TIR)-encoding domain of the N transcript (Li et al., 2012; Moon and Park, 2016). Moreover, synthesis of phased, secondary siRNAs (phasiRNAs) from the N coding sequence through overexpression of miR6019 was shown to be accompanied by reduction in the N gene transcript accumulation and N-mediated resistance against TMV (Li et al., 2012).

In present study, the expression of miRNAs was potentially regulated by PRSV infection in papaya plant lines that were characterized by using high-throughput sequencing technology. The molecular functions and validation of these miRNAs in inducing resistance against PRSV infection in the resistant plants requires further investigation. However, the results presented here will improve the understanding of the viral infection and the response of miRNAs in the host plants in general and in particular for PRSV in papaya plants. In the future, this knowledge will also be helpful in providing additional information for developing viral management strategies, mainly through the plant breeding technologies and biotechnological interventions.

Data availability statement

The original contributions presented in the study are publicly available. This data can be found here: NCBI SRA, PRJNA1122089.

Author contributions

BP: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. ST: Resources, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. BP acknowledges Research funding from DBT and ICAR (Govt. of India).

Acknowledgments

Technical help by Dr. Shivaraj M in carrying out the RT-PCRs is acknowledged. Authors also acknowledge Dr. SK Sharma and Dr. Raj Verma for sharing the pictures of papaya field trial.

Conflict of interest

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

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2024.1398437/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Workflow for the Illumina (HiSeq 2500) next generation sequencing data analysis.

SUPPLEMENTARY TABLE 1
Primers used for screening for PRSV and in the validation of NGS data by RT-PCR.

SUPPLEMENTARY TABLE 2
List of miRNAs upregulated/downregulated and the corresponding genes downregulated/upregulated along with the description of the putative target genes in Papaya.

References

Abreu, P. M., Gaspar, C. G., Buss, D. S., Ventura, J. A., Ferreira, P. C. G., and Fernandes, P. M. B. (2014). Carica papaya MicroRNAs Are Responsive to Papaya meleira virus Infection. *PloS One* 9, e103401. doi: 10.1371/journal.pone.0103401

Al Abdallat, A. M., Ayad, J. Y., Abu Elenein, J. M., al Ajlouni, Z., and Harwood, W. A. (2014). Overexpression of the transcription factor HvSNAC1 improves drought tolerance in barley (Hordeum vulgare L.). *Mol. Breed.* 33, 401–414. doi: 10.1007/s11032-013-9958-1

Ambros, V. (2001). microRNAs: tiny regulators with great potential. *Cell* 107, 823–826. doi: 10.1016/S0092-8674(01)00616-X

Andrés-León, E., Núñez-Torres, R., and Rojas, A. M. (2016). miARma-Seq: a comprehensive tool for miRNA, mRNA and circRNA analysis. *Sci. Rep.* 6, 25749. doi: 10.1038/srep25749

Axtell, M. J., and Bartel, D. P. (2005). Antiquity of microRNAs and their targets in land plants. *Plant Cell* 17, 1658–1673. doi: 10.1105/tpc.105.032185

Axtell, M. J., Snyder, J. A., and Bartel, D. P. (2007). Common functions for diverse small RNAs of land plants. *Plant Cell* 19, 1750–1769. doi: 10.1105/tpc.107.051706

Bao, D., Ganbaatar, O., Cui, X., Yu, R., Bao, W., Falk, B. W., et al. (2018). Down-regulation of genes coding for core RNAi components and disease resistance proteins via corresponding microRNAs might be correlated with successful Soybean mosaic virus infection in soybean. *Mol. Plant Pathol.* 19, 948–960. doi: 10.1111/mpp.12581

Bartel, D. P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116, 281–297. doi: 10.1016/S0092-8674(04)00045-5

Bazzini, A. A., Hopp, H. E., Beachy, R. N., and Asurmendi, S. (2007). Infection and coaccumulation of tobacco mosaic virus proteins alter microRNA levels, correlating with symptom and plant development. *Proc. Natl. Acad. Sci.* 104, 12157–12162. doi: 10.1073/pnas.0705114104

Ben-Gera, H., Dafna, A., Alvarez, J. P., Bar, M., Mauerer, M., and Ori, N. (2016). Auxin-mediated lamina growth in tomato leaves is restricted by two parallel mechanisms. *Plant J.* 86, 443–457. doi: 10.1111/tpj.13188

Bohmert, K., Camus, I., Bellini, C., Bouchez, D., Caboche, M., and Benning, C. (1998). AGO1 defines a novel locus of Arabidopsis controlling leaf development. *EMBO J.* 17, 170–180. doi: 10.1093/emboj/17.1.170

Boyko, A., Kathiria, P., Zemp, F. J., Yao, Y., Pogribny, I., and Kovalchuk, I. (2007). Transgenerational changes in the genome stability and methylation in pathogen-infected plants: (virus-induced plant genome instability). *Nucleic Acids Res.* 35, 1714–1725. doi: 10.1093/nar/gkm029

Bresso, E. G., Chorostecki, U., Rodriguez, R. E., Palatnik, J. F., and Schommer, C. (2018). Spatial control of gene expression by miR319-regulated TCP transcription factors in leaf development. *Plant Physiol.* 176, 1694–1708. doi: 10.1104/pp.17.00823

Calil, I. P., and Fontes, E. P. B. (2017). Plant immunity against viruses: antiviral immune receptors in focus. *Ann. Bot.* 119 (5), 711–723. doi: 10.1093/aob/mcw200

Carrington, J. C., and Ambros, V. (2003). Role of microRNAs in plant and animal development. *Sci. (New York N.Y.)* 301, 336–338. doi: 10.1126/science.1085242

Chapman, E. J., Prokhnevsky, A. I., Gopinath, K., Dolja, V. V., and Carrington, J. C. (2004). Viral RNA silencing suppressors inhibit the microRNA pathway at an intermediate step. *Genes Dev.* 18, 1179–1186. doi: 10.1101/gad.1201204

Chavan, V. M., Tomar, S. P. S., and Dhale, M. G. (2010). Management of papaya ring spot virus (prsv-p) of papaya under pune conditions. *Acta Hortic.* 851, 447–452. doi: 10.17660/ActaHortic.2010.851.69

Chen, X. (2005). MicroRNA biogenesis and function in plants. *FEBS Lett.* 579, 5923–5931. doi: 10.1016/j.febslet.2005.07.071

Chen, H., Zhang, L., Yu, K., and Wang, A. (2015). Pathogenesis of Soybean mosaic virus in soybean carrying Rsv1 gene is associated with miRNA and siRNA pathways, and breakdown of AGO1 homeostasis. *Virology* 476, 395–404. doi: 10.1016/j.virol.2014.12.034

Chen, H., Arsovski, A. A., Yu, K., and Wang, A. (2016). Genome-Wide Investigation Using sRNA-Seq, Degradome-Seq and Transcriptome-Seq Reveals Regulatory Networks of microRNAs and Their Target Genes in Soybean during Soybean mosaic virus Infection. *PloS One* 11, e0150582. doi: 10.1371/journal.pone.0150582

Chiou, T. (2007). The role of microRNAs in sensing nutrient stress. *Plant Cell Environ.* 30, 323–332. doi: 10.1111/j.1365-3040.2007.01643.x

Cui, X., Yan, Q., Gan, S., Xue, D., Dou, D., Guo, N., et al. (2017). Overexpression of gma-miR1510a/b suppresses the expression of a NB-LRR domain gene and reduces resistance to Phytophthora sojae. *Gene* 621, 32–39. doi: 10.1016/j.gene.2017.04.015

Dai, X., Zhuang, Z., and Zhao, P. X. (2018). psRNATarget: a plant small RNA target analysis server, (2017 release). *Nucleic Acids Res.* 46, W49–W54. doi: 10.1093/nar/gky316

Datar, W., Zote, K., Verma, R., Tripathi, S., and Tomer, S. (2013). Field evaluation of papaya cultivars and lines for reaction to Papaya ring spot virus. *J. Mycol. Plant Pathol.* 43, 481–483.

Ding, S.-W., and Lu, R. (2011). Virus-derived siRNAs and piRNAs in immunity and pathogenesis. *Curr. Opin. Virol.* 1, 533–544. doi: 10.1016/j.coviro.2011.10.028

Du, J., Wu, G., Zhou, Z., Zhang, J., Li, M., Sun, M., et al. (2019). Identification of microRNAs regulated by tobacco curly shoot virus co-infection with its betasatellite in Nicotiana benthamiana. *Virol. J.* 16, 130. doi: 10.1186/s12985-019-1234-5

Fahlgren, N., Howell, M. D., Kasschau, K. D., Chapman, E. J., Sullivan, C. M., Cumbie, J. S., et al. (2007). High-throughput sequencing of arabidopsis microRNAs: evidence for frequent birth and death of MIRNA genes. *PloS One* 2, e219. doi: 10.1371/journal.pone.0000219

Feng, H., Duan, X., Zhang, Q., Li, X., Wang, B., Huang, L., et al. (2014). The target gene of tae-miR164, a novel NAC transcription factor from the NAMsubfamily, negatively regulates resistance of wheat to stripe rust. *Mol. Plant Pathol.* 15, 284–296. doi: 10.1111/mpp.12089

Garcia, D., Collier, S. A., Byrne, M. E., and Martienssen, R. A. (2006). Specification of leaf polarity in Arabidopsis via the trans-acting siRNA pathway. *Curr. Biol.: CB* 16, 933–938. doi: 10.1016/j.cub.2006.03.064

Gonsalves, D., Tripathi, S., Carr, J. B., and Suzuki, J. Y. (2010). *Papaya ringspot virus* (St. Paul, MN 55121 USA: The American Phytopathological Society (APS)). doi: 10.1094/PHI-I-2010-1004-01

Gorane, A., Verma, R., Naik, A., Nikam, T., Ade, A., Mahapatro, G., et al. (2019). Characterization of a severe isolate of papaya ringspot virus from papaya in western India. *J. Plant Pathol.* 101, 1203–1209. doi: 10.1007/s42161-019-00340-4

Gorane, A., Verma, R., Naik, A., Nikam, T., Ade, A., and Tripathi, S. (2021). Genetic variability and recombination analysis of papaya infecting isolates of Papaya ringspot virus based on CP gene from Western India. *Arch. Phytopathol. Plant Prot.* 54, 2294–2305. doi: 10.1080/03235408.2021.1975463

Grimson, A., Srivastava, M., Fahey, B., Woodcroft, B. J., Chiang, H. R., King, N., et al. (2008). Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. *Nature* 455, 1193–1197. doi: 10.1038/nature07415

Guo, H.-S., Xie, Q., Fei, J.-F., and Chua, N.-H. (2005). MicroRNA directs mRNA cleavage of the transcription factor NAC1 to downregulate auxin signals for arabidopsis lateral root development. *Plant Cell* 17, 1376–1386. doi: 10.1105/tpc.105.030841

Husbands, A. Y., Benkovics, A. H., Nogueira, F. T. S., Lodha, M., and Timmermans, M. C. P. (2016). The ASYMMETRIC LEAVES complex employs multiple modes of regulation to affect adaxial-abaxial patterning and leaf complexity. *Plant Cell* 27 (12), 3321–3335. doi: 10.1105/tpc.15.00454

Jia, X., Ding, N., Fan, W., Yan, J., Gu, Y., Tang, X., et al. (2015). Functional plasticity of miR165/166 in plant development revealed by small tandem target mimic. *Plant Sci.: Int. J. Exp. Plant Biol.* 233, 11–21. doi: 10.1016/j.plantsci.2014.12.020

Jones-Rhoades, M. W., and Bartel, D. P. (2004). Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. *Mol. Cell* 14, 787–799. doi: 10.1016/j.molcel.2004.05.027

Jones-Rhoades, M. W., Bartel, D. P., and Bartel, B. (2006). MicroRNAS and their regulatory roles in plants. *Annu. Rev. Plant Biol.* 57, 19–53. doi: 10.1146/annurev.arplant.57.032905.105218

Kang, M., Zhao, Q., Zhu, D., and Yu, J. (2012). Characterization of microRNAs expression during maize seed development. *BMC Genomics* 13, 360. doi: 10.1186/1471-2164-13-360

Khalid, A., Zhang, Q., Yasir, M., and Li, F. (2017). Small RNA based genetic engineering for plant viral resistance: application in crop protection. *Front. Microbiol.* 8. doi: 10.3389/fmicb.2017.00043

Khraiwesh, B., Zhu, J.-K., and Zhu, J. (2012). Role of miRNAs and siRNAs in biotic and abiotic stress responses of plants. *Biochim. Biophys. Acta* 1819, 137–148. doi: 10.1016/j.bbagrm.2011.05.001

Kovalchuk, I., Kovalchuk, O., Kalck, V., Boyko, V., Filkowski, J., Heinlein, M., et al. (2003). Pathogen-induced systemic plant signal triggers DNA rearrangements. *Nature* 423, 760–762. doi: 10.1038/nature01683

Kumar, R. (2014). Role of microRNAs in biotic and abiotic stress responses in crop plants. *Appl. Biochem. Biotechnol.* 174, 93–115. doi: 10.1007/s12010-014-0914-2

Kundu, A., Paul, S., Dey, A., and Pal, A. (2017). High throughput sequencing reveals modulation of microRNAs in Vigna mungo upon Mungbean Yellow Mosaic India Virus inoculation highlighting stress regulation. *Plant Sci.* 257, 96–105. doi: 10.1016/j.plantsci.2017.01.016

Kurihara, Y., Takashi, Y., and Watanabe, Y. (2006). The interaction between DCL1 and HYL1 is important for efficient and precise processing of pri-miRNA in plant microRNA biogenesis. *RNA* 12, 206–212. doi: 10.1261/rna.2146906

Lauter, N., Kampani, A., Carlson, S., Goebel, M., and Moose, S. P. (2005). microRNA172 down-regulates glossy15 to promote vegetative phase change in maize. *Proc. Natl. Acad. Sci. U.S.A.* 102, 9412–9417. doi: 10.1073/pnas.0503927102

Li, S., Liu, W., Zhang, X., Liu, Y., Li, N., and Li, Y. (2012). Roles of the Arabidopsis G protein γ subunit AGG3 and its rice homologs GS3 and DEP1 in seed and organ size control. *Plant Signal. Behav.* 7, 1357–1359. doi: 10.4161/psb.21620

Li, Y., Zhao, S.-L., Li, J.-L., Hu, X.-H., Wang, H., Cao, X.-L., et al. (2017). Osa-miR169 Negatively Regulates Rice Immunity against the Blast Fungus Magnaporthe oryzae. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.00002

Liang, G., Li, Y., He, H., Wang, F., and Yu, D. (2013). Identification of miRNAs and miRNA-mediated regulatory pathways in Carica papaya. *Planta* 238, 739–752. doi: 10.1007/s00425-013-1929-6

Liang, C., Liu, H., Hao, J., Li, J., and Luo, L. (2019). Expression profiling and regulatory network of cucumber microRNAs and their putative target genes in response to cucumber green mottle mosaic virus infection. *Arch. Virol.* 164, 1121–1134. doi: 10.1007/s00705-019-04152-w

Liu, Z., Jia, L., Wang, H., and He, Y. (2011). HYL1 regulates the balance between adaxial and abaxial identity for leaf flattening via miRNA-mediated pathways. *J. Exp. Bot.* 62, 4367–4381. doi: 10.1093/jxb/err167

Liu, K., Yuan, C., Li, H., Lin, W., Yang, Y., Shen, C., et al. (2015a). Genome-wide identification and characterization of auxin response factor (ARF) family genes related to flower and fruit development in papaya (*Carica papaya* L.). *BMC Genomics* 16, 901. doi: 10.1186/s12864-015-2182-0

Liu, S.-R., Zhou, J.-J., Hu, C.-G., Wei, C.-L., and Zhang, J.-Z. (2017). MicroRNA-mediated gene silencing in plant defense and viral counter-defense. *Front. Microbiol.* 8. doi: 10.3389/fmicb.2017.01801

Loebenstein, G. (2009). Local lesions and induced resistance. *Adv. Virus Res.* 75, 73–117. doi: 10.1016/S0065-3527(09)07503-4

Luan, Y., Cui, J., Zhai, J., Li, J., Han, L., and Meng, J. (2015). High-throughput sequencing reveals differential expression of miRNAs in tomato inoculated with Phytophthora infestans. *Planta* 241, 1405–1416. doi: 10.1007/s00425-015-2267-7

Marino, D., Dunand, C., Puppo, A., and Pauly, N. (2012). A burst of plant NADPH oxidases. *Trends Plant Sci.* 17, 9–15. doi: 10.1016/j.tplants.2011.10.001

Martins, T. F., Souza, P. F. N., Alves, M. S., Silva, F. D. A., Arantes, M. R., Vasconcelos, I. M., et al. (2020). Identification, characterization, and expression analysis of cowpea (*Vigna unguiculata* [L.] Walp.) miRNAs in response to cowpea severe mosaic virus (CPSMV) challenge. *Plant Cell Rep.* 39, 1061–1078. doi: 10.1007/s00299-020-02548-6

Mishra, R., Gaur, R. K., and Patil, B. L. (2016). "Current knowledge of viruses infecting papaya and their transgenic management," in *Plant viruses, evolution and management*. Eds. R. Gaur, N. Petrov, B. Patil and M. Stoyanova (Springer, Singapore), 189–203.

Moon, J. Y., and Park, J. M. (2016). Cross-talk in viral defense signalling in plants. *Front. Microbiol.* 07. doi: 10.3389/fmicb.2016.02068

Moxon, S., Jing, R., Szittya, G., Schwach, F., Rusholme Pilcher, R. L., Moulton, V., et al. (2008). Deep sequencing of tomato short RNAs identifies microRNAs targeting genes involved in fruit ripening. *Genome Res.* 18, 1602–1609. doi: 10.1101/gr.080127.108

Muthamilarasan, M., and Prasad, M. (2013). Plant innate immunity: an updated insight into defense mechanism. *J. Biosci.* 38, 433–449. doi: 10.1007/s12038-013-9302-2

Naqvi, A. R., Haq, Q. M., and Mukherjee, S. K. (2010). MicroRNA profiling of tomato leaf curl new Delhi virus (ToLCNDV) infected tomato leaves indicates that deregulation of mir159/319 and mir172 might be linked with leaf curl disease. *Virol. J.* 7, 281. doi: 10.1186/1743-422X-7-281

Omidbakhshfard, M. A., Proost, S., Fujikura, U., and Mueller-Roeber, B. (2015). Growth-regulating factors (GRFs): A small transcription factor family with important functions in plant biology. *Mol. Plant* 8, 998–1010. doi: 10.1016/j.molp.2015.01.013

Pacheco, R., García-Marcos, A., Barajas, D., Martiáñez, J., and Tenllado, F. (2012). PVX–potyvirus synergistic infections differentially alter microRNA accumulation in *Nicotiana benthamiana*. *Virus Res.* 165, 231–235. doi: 10.1016/j.virusres.2012.02.012

Parameswari, B., Mangrauthia, S. K., Praveen, S., and Jain, R. K. (2007). Complete genome sequence of an isolate of Papaya ringspot virus from India. *Arch. Virol.* 152, 843–845. doi: 10.1007/s00705-006-0890-0

Park, M. R., Seo, J. K., and Kim, K. H. (2013). Viral and nonviral elements in potexvirus replication and movement and in antiviral responses. *Adv. Virus Res.* 87, 75–112. doi: 10.1016/B978-0-12-407698-3.00003-X

Patil, B. L., Chakraborty, S., Czosnek, H., Fiallo-Olivé, E., Gilbertson, R. L., Legg, J., et al. (2021). "Plant Resistance to Geminiviruses," in *Encyclopedia of Virology* (Oxford: Academic Press), 554–566. doi: 10.1016/B978–0-12–809633–8.21565–3

Patwa, N., Nithin, C., Bahadur, R. P., and Basak, J. (2019). Identification and characterization of differentially expressed *Phaseolus vulgaris* miRNAs and their targets during mungbean yellow mosaic India virus infection reveals new insight into Phaseolus-MYMIV interaction. *Genomics* 111, 1333–1342. doi: 10.1016/j.ygeno.2018.09.005

Peng, T., Sun, H., Qiao, M., Zhao, Y., Du, Y., Zhang, J., et al. (2014). Differentially expressed microRNA cohorts in seed development may contribute to poor grain filling of inferior spikelets in rice. *BMC Plant Biol.* 14, 196. doi: 10.1186/s12870-014-0196-4

Prakash, J., and Singh, A. K. (2013). Screening of Papaya genotypes against viral diseases. *Indian J. Horticult.* 70, 437–438.

Prigge, M. J., and Wagner, D. R. (2001). The Arabidopsis serrate gene encodes a zinc-finger protein required for normal shoot development. *Plant Cell* 13, 1263–1279. doi: 10.1105/TPC.010095

Rhoades, M. W., Reinhart, B. J., Lim, L. P., Burge, C. B., Bartel, B., and Bartel, D. P. (2002). Prediction of plant microRNA targets. *Cell* 110, 513–520. doi: 10.1016/S0092-8674(02)00863-2

Sett, S., Prasad, A., and Prasad, M. (2022). Resistance genes on the verge of plant-virus interaction. *Trends Plant Sci.* 27, 1242–1252. doi: 10.1016/j.tplants.2022.07.003

Sharma, S. K., and Tripathi, S. (2019). Horticultural characterization and papaya ringspot virus reaction of papaya Pune Selections. *Indian J. Horticult.* 76, 32. doi: 10.5958/0974-0112.2019.00005.7

Sharma, S. K., Tripathi, S., and Mitra, S. K. (2019). Breeding for resistance against Papaya ringspot virus: history, present status and future prospects in India. *Acta Hortic.* 1250, 45–54. doi: 10.17660/ActaHortic.2019.1250.8

Shivaprasad, P., Chen, H.-M., Patel, K., Bond, D. M., Santos, B. A. C. M., and Baulcombe, D. C. (2012). A microRNA superfamily regulates nucleotide binding site–leucine-rich repeats and other mRNAs. *Plant Cell* 24, 859–874. doi: 10.1105/tpc.111.095380

Singh, A. K., Bajpai, A., and Singh, A. (2006). Assessment of genetic diversity in papaya (*Carica papaya*) cultivars in northern plains. *Indian J. Agric. Sci.* 76, 692–694.

Soltani, N., Best, T., Grace, D., Nelms, C., Shumaker, K., Romero-Severson, J., et al. (2020). Transcriptome profiles of Quercus rubra responding to increased O3 stress. *BMC Genomics* 21, 160. doi: 10.1186/s12864-020-6549-5

Soltani, N., Staton, M., and Gwinn, K. D. (2021). Response of bitter and sweet Chenopodium quinoa varieties to cucumber mosaic virus: Transcriptome and small RNASeq perspective. *PloS One* 16, e0244364. doi: 10.1371/journal.pone.0244364

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 2017 update. *Nucleic Acids Res.* 45, W122–W129. doi: 10.1093/nar/gkx382

Tiwari, M., Sharma, D., and Trivedi, P. K. (2014). Artificial microRNA mediated gene silencing in plants: progress and perspectives. *Plant Mol. Biol.* 86, 1–18. doi: 10.1007/s11103-014-0224-7

Tripathi, S., Suzuki, J. Y., Ferreira, S. A., and Gonsalves, D. (2008). Papaya ringspot virus-P: characteristics, pathogenicity, sequence variability and control. *Mol. Plant Pathol.* 9, 269–280. doi: 10.1111/j.1364-3703.2008.00467.x

Várallyay, É., Válóczi, A., Ágyi, Á., Burgyán, J., and Havelda, Z. (2010). Plant virus-mediated induction of miR168 is associated with repression of ARGONAUTE1 accumulation. *EMBO J.* 29, 3507–3519. doi: 10.1038/emboj.2010.215

Weiberg, A., and Jin, H. (2015). Small RNAs — the secret agents in the plant–pathogen interactions. *Curr. Opin. Plant Biol.* 26, 87–94. doi: 10.1016/j.pbi.2015.05.033

Whitham, S., Dinesh-Kumar, S. P., Choi, D., Hehl, R., Corr, C., and Baker, B. (1994). The product of the tobacco mosaic virus resistance gene N: Similarity to toll and the interleukin-1 receptor. *Cell* 78, 1101–1115. doi: 10.1016/0092-8674(94)90283-6

Williams, L., Grigg, S. P., Xie, M., Christensen, S., and Fletcher, J. C. (2005). Regulation of Arabidopsis shoot apical meristem and lateral organ formation by microRNA miR166g and its AtHD-ZIP target genes. *Development* 132, 3657–3668. doi: 10.1242/dev.01942

Winter, A. D., Weir, W., Hunt, M., Berriman, M., Gilleard, J. S., Devaney, E., et al. (2012). Diversity in parasitic nematode genomes: the microRNAs of Brugia pahangi and Haemonchus contortus are largely novel. *BMC Genomics* 13, 4. doi: 10.1186/1471-2164-13-4

Xiao, B., Yang, X., Ye, C.-Y., Liu, Y., Yan, C., Wang, Y., et al. (2014). A diverse set of miRNAs responsive to begomovirus-associated betasatellite in Nicotiana benthamiana. *BMC Plant Biol.* 14, 60. doi: 10.1186/1471-2229-14-60

Xin, M., Wang, Y., Yao, Y., Xie, C., Peng, H., Ni, Z., et al. (2010). Diverse set of microRNAs are responsive to powdery mildew infection and heat stress in wheat (Triticum aestivum L.). *BMC Plant Biol.* 10, 123. doi: 10.1186/1471-2229-10-123

Xu, H., He, X., Zheng, X., Yang, Y., Tian, J., and Lu, Z. (2014). Infection of rice plants by rice black streaked dwarf virus improves an egg parasitoid, *anagrus nilaparvatae* (Hymenoptera: mymaridae), of rice planthoppers. *Environ. Entomol.* 43, 1235–1239. doi: 10.1603/EN14044

Xu, M., Hu, T., Zhao, J., Park, M.-Y., Earley, K. W., Wu, G., et al. (2016). Developmental Functions of miR156-Regulated SQUAMOSA PROMOTER BINDING PROTEIN-LIKE (SPL) Genes in Arabidopsis thaliana. *PloS Genet.* 12, e1006263. doi: 10.1371/journal.pgen.1006263

Yin, X., Wang, J., Cheng, H., Wang, X., and Yu, D. (2013). Detection and evolutionary analysis of soybean miRNAs responsive to soybean mosaic virus. *Planta* 237, 1213–1225. doi: 10.1007/s00425-012-1835-3

Zhai, J., Jeong, D.-H., de Paoli, E., Park, S., Rosen, B. D., Li, Y., et al. (2011). MicroRNAs as master regulators of the plant NB-LRR defense gene family via the production of phased, trans-acting siRNAs. *Genes Dev.* 25, 2540–2553. doi: 10.1101/gad.177527.111

Zhang, Y. (2005). miRU: an automated plant miRNA target prediction server. *Nucleic Acids Res.* 33, W701–W704. doi: 10.1093/nar/gki383

Zheng, Z., Wang, N., Jalajakumari, M., Blackman, L., Shen, E., Verma, S., et al. (2020). miR159 represses a constitutive pathogen defense response in tobacco. *Plant Physiol.* 182, 2182–2198. doi: 10.1104/pp.19.00786

Zhu, H., Zhou, Y., Castillo-González, C., Lu, A., Ge, C., Zhao, Y.-T., et al. (2013). Bidirectional processing of pri-miRNAs with branched terminal loops by Arabidopsis Dicer-like1. *Nat. Struct. Mol. Biol.* 20, 1106–1115. doi: 10.1038/nsmb.2646

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 03 July 2024
DOI 10.3389/fpls.2024.1386494

Check for updates

OPEN ACCESS

EDITED BY
Brigitte Mauch-Mani,
Université de Neuchâtel, Switzerland

REVIEWED BY
Jindong Liu,
Chinese Academy of Agricultural Sciences, China
Arun Kumar Joshi,
International Maize and Wheat Improvement Center (South Asia), India

*CORRESPONDENCE
Sundeep Kumar
Sundeep.kumar@icar.gov.in
Neeraj Budhlakoti
neeraj.budhlakoti@icar.gov.in

†PRESENT ADDRESS
Dinesh Kumar Saini,
Department of Plant and Soil Science, Texas Tech University, Lubbock, TX, United States

RECEIVED 15 February 2024
ACCEPTED 11 June 2024
PUBLISHED 03 July 2024

CITATION
Sharma D, Budhlakoti N, Kumari A, Saini DK, Sharma A, Yadav A, Mir RR, Singh AK, Vikas VK, Singh GP and Kumar S (2024) Exploring the genetic architecture of powdery mildew resistance in wheat through QTL meta-analysis.
Front. Plant Sci. 15:1386494.
doi: 10.3389/fpls.2024.1386494

COPYRIGHT
© 2024 Sharma, Budhlakoti, Kumari, Saini, Sharma, Yadav, Mir, Singh, Vikas, Singh and Kumar. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Exploring the genetic architecture of powdery mildew resistance in wheat through QTL meta-analysis

Divya Sharma[1], Neeraj Budhlakoti[2*], Anita Kumari[3], Dinesh Kumar Saini[4†], Anshu Sharma[1], Aakash Yadav[1], Reyazul Rouf Mir[5], Amit Kumar Singh[1], V. K. Vikas[6], Gyanendra Pratap Singh[1] and Sundeep Kumar[1*]

[1]Divison of Genomic Resources, ICAR-National Bureau of Plant Genetic Resources, New Delhi, India, [2]Centre for Agriculture Bioinformatics, ICAR-Indian Agricultural Statistics Research Institute, New Delhi, India, [3]Department of Botany, University of Delhi, Delhi, India, [4]Department of Plant Breeding and Genetics, Punjab Agricultural University, Punjab, Ludhiana, India, [5]Department of Genetics and Plant Breeding , Sher-e-Kashmir University of Agricultural Sciences & Technology of Kashmir (SKUAST-K), Srinagar, Kashmir, India, [6]Divison of Crop Improvement, ICAR-Indian Agricultural Research Institute, Regional Station, Wellington, Tamilnadu, India

Powdery mildew (PM), caused by *Blumeria graminis f.* sp. *tritici*, poses a significant threat to wheat production, necessitating the development of genetically resistant varieties for long-term control. Therefore, exploring genetic architecture of PM in wheat to uncover important genomic regions is an important area of wheat research. In recent years, the utilization of meta-QTL (MQTL) analysis has gained prominence as an essential tool for unraveling the complex genetic architecture underlying complex quantitative traits. The aim of this research was to conduct a QTL meta-analysis to pinpoint the specific genomic regions in wheat responsible for governing PM resistance. This study integrated 222 QTLs from 33 linkage-based studies using a consensus map with 54,672 markers. The analysis revealed 39 MQTLs, refined to 9 high-confidence MQTLs (hcMQTLs) with confidence intervals of 0.49 to 12.94 cM. The MQTLs had an average physical interval of 41.00 Mb, ranging from 0.000048 Mb to 380.71 Mb per MQTL. Importantly, 18 MQTLs co-localized with known resistance genes like *Pm2*, *Pm3*, *Pm8*, *Pm21*, *Pm38*, and *Pm41*. The study identified 256 gene models within hcMQTLs, providing potential targets for marker-assisted breeding and genomic prediction programs to enhance PM resistance. These MQTLs would serve as a foundation for fine mapping, gene isolation, and functional genomics studies, facilitating a deeper understanding of molecular mechanisms. The identification of candidate genes opens up exciting possibilities for the development of PM-resistant wheat varieties after validation.

KEYWORDS

wheat, mQTL, candidate gene, confidence interval, GWAS, consensus map

Introduction

Powdery mildew (PM), caused by the obligate biotrophic fungus *Blumeria graminis* s *f.* sp. *tritici (Bgt)*, is a widespread global disease resulting in substantial wheat yield losses. Over the past forty years, there have been numerous global outbreaks of wheat PM (Morgounov et al., 2012). The pathogen is the sixth most significant fungal disease of wheat (Dean et al., 2012) and the eighth largest cause of wheat production loss worldwide (Savary et al., 2019). Generally, PM has predominantly been observed in regions characterized by ample rainfall and humid climate (Bennett, 1984). Nonetheless, its emergence as a significant concern in drier and warmer regions is leading to substantial reductions in wheat yield in these areas. This shift could be attributed to the cultivation of semi-dwarf wheat varieties, the increased use of nitrogen-based fertilizers, and the practice of dense planting in the field (Wang et al., 2005; Morgounov et al., 2012). Resistance towards PM is genetically inherited in wheat crop and is controlled by both race-specific and non-race specific genes conferring a qualitative and quantitative resistance. Race-specific resistance is largely heritable, conferred by a single resistance gene, and provides complete resistance to some specific pathogen infections but not others. Non-racial resistance provides a form of partial resistance that is not reliant on specific pathogen avirulence genes. This type of resistance permits infection but limits the spread of the disease (Maurya et al., 2021; Cheng et al., 2022; Mapuranga et al., 2022c).

Most PM resistance research has concentrated on key genes that are thought to be qualitatively race-specific. For instance, extensive research is conducted on the *Pm3* resistance gene and its various alleles due to their ease of manipulation, transient expression, simple inheritance, and ability to confer complete resistance. This resistance is often associated with the hypersensitive response, but it is effective only against a small number of pathogen races and can be readily overcome by new virulent pathogen strains (Yahiaoui et al., 2004; Koller et al., 2018; Simeone et al., 2020). Until recently, the primary emphasis in plant studies has been on adult-plant resistance, which is associated with long-lasting and non-specific resistance. This resistance involves the interplay of multiple genes that delay and obstruct fungal infection, growth, and reproduction during the adult-plant phase (Jakobson et al., 2012). Therefore, combining multiple resistance genes is considered as the most economical and environmentally friendly strategy for increasing the persistence of resistance against most fungal infections in wheat. The attainment of this goal necessitates the integration of conventional breeding methods with molecular techniques, offering the potential to enhance the selection efficiency for resistance to PM and traits related to crop yield (Bapela et al., 2023).In recent decades, the introduction of Next-Generation Sequencing (NGS) technology for genotyping and advancements in genomic-assisted breeding has significantly expedited the identification and incorporation of genes for PM resistance into commercially grown wheat varieties. Moreover, 19 PM resistance genes/alleles have been cloned, e.g., *Pm1a*, *Pm2*, *Pm3* (*Pm3a to Pm3g, Pm3k*, and *Pm3r*), *Pm4*, *Pm8*, *Pm17*, *Pm21*, *Pm24*, *Pm38/Yr18/Lr34/Sr57*, *Pm46/Yr46/Lr67/Sr55*, *Pm60* and *WTK4* (Yahiaoui et al., 2004; Srichumpa et al., 2005; Bhullar et al., 2009; Cao et al., 2011; Hurni et al., 2014; He et al., 2018; Hewitt et al., 2021; Sánchez-Martín et al., 2021; Gaurav et al., 2022). Of these, only *Pm3k* belongs to tetraploid wheat (Yahiaoui et al., 2009). Most of these genes have been incorporated into wheat from closely related species and their wild relatives. However, due to their reduced resistance levels and the undesirable linkage drag they carry, these genes have not seen widespread commercial utilization (Friebe et al., 1994).

Various research groups have previously developed high-density linkage maps and used them in QTL mapping studies for PM resistance (Singh et al., 2022). Nonetheless, the majority of the QTLs identified in different research studies have not undergone fine mapping due to overlapping genomic regions and have seldom been utilized in marker-assisted breeding programs. This appears to be due to variations in experimental designs, environmental conditions, genetic backgrounds of the parental strains, population sizes, types, and densities of molecular markers employed, as well as the statistical methodologies employed in subsequent analyses. Moreover, when focusing on wheat specifically, additional factors come into play, including the complexity of the hexaploid wheat genome, the prevalence of highly repetitive sequences within the genome, and the lack of comprehensive high-density linkage maps (Kumar et al., 2023).

Meta-analysis of QTLs retrieved from different independent studies, is an alternate method that can help in precise mapping of traits (Sharma et al., 2023). MQTL analysis is a relatively new concept and is rapidly emerging an efficient method for narrowing the confidence intervals (CI) of overlapping QTLs, allowing for rapid and efficient discovery of candidate markers and genes linked to the trait of interest (Kumari et al., 2023; Sharma et al., 2023). Meta-analysis has already been performed for various traits in wheat (Kumar et al., 2021; Pal et al., 2021, Saini et al., 2022b) including resistance to different diseases such as leaf rust (Soriano and Royo, 2015; Amo and Soriano, 2022), stem rust (Yu et al., 2014), tan spot (Liu et al., 2020), Fusarium head blight (Liu et al., 2009; Löffler et al., 2009; Venske et al., 2019; Zheng et al., 2021), stripe rust (Jan et al., 2021; Kumar et al., 2023), multiple disease resistance (Pal et al., 2022; Saini et al., 2022a) and PM resistance (Marone et al., 2013). The most recent study on PM resistance, published in 2013, conducted a meta-analysis with a dataset consisting of only 96 QTLs and identified 24 MQTLs (Marone et al., 2013). Since considerable number of QTLs for parameters contributing to PM resistance have been reported after this last report of MQTL analysis for PM resistance in wheat, the present study involving MQTL analysis was conducted (based on QTL studies published until July 2021) to supplement the list of MQTLs and candidate genes reported in the earlier MQTL study for PM resistance. Overall, the primary aim of this study is to investigate the genetic basis of PM resistance by identifying promising genomic regions and candidate genes using the newly available wheat genome data, integrating it with GWAS information and analyzing the roles of the identified candidate genes in various wheat tissues. The outcomes of this research will have significant utility for wheat breeders, providing valuable resources for enhancing resistance to PM in wheat varieties.

Materials and methods

Literature survey for collection of QTLs for powdery mildew

We conducted a comprehensive search for articles that reported QTLs linked to PM resistance in wheat, spanning from 1996 to 2021. This search was performed using Google Scholar (https://scholar.google.com/) and other accessible data sources. An additional resource for this search was a recently created Wheat QTL database (http://wheatqtldb.net/; Singh et al., 2022). Each QTL mapping study was screened to extract the following information: (i) mapping population type (e.g., $F_{2:3}$, RILs, DH, BC) and their parents, (ii) population size, (iii) chromosome number, (iv) position of the QTLs and flanking markers, (v) logarithm of odds (LOD) values and (vi) variance explained by the individual QTLs (PVE) and (vii) different disease resistance parameters, such as area under disease progression cure (AUDPC), infection type (IT), maximum disease severity (MDS) and vernalized seedling score (VSS). QTLs with missing data were excluded from the meta-analysis.

Development of a comprehensive consensus map

The R package called 'LPmerge' (Endelman and Plomion, 2014) was used to construct the consensus map following the steps described recently (Kumar et al., 2023). The following genetic maps were employed as the framework maps in this process: (i) the 'ITMI_SSR map,' containing 1406 loci (Song et al., 2005); (ii) the 'Wheat, Consensus SSR, 2004' map, which includes 1235 marker loci (Somers et al., 2004); (iii) an integrated map for durum wheat with 30,144 markers (Maccaferri et al., 2015); and (iv) the 'Illumina iSelect 90 K SNP Array-based genetic map,' comprising 40,267 loci (Wen et al., 2017). For the construction of the consensus map, markers flanking individual QTLs were also included. The genetic maps contained several common markers located at different genetic positions, and these markers were taken into consideration during the construction of the consensus chromosomes maps.

QTL projection on the consensus map

The genetic map file and the QTL information file from each study were compiled and employed as input text files for QTL projection through BioMercator V4.2 (Sosnowski et al., 2012). This software requires a set of distinct descriptors for each QTL such as the genetic position of the QTL (both peak position and CI), LOD score, PVE value, the trait linked with the QTL, and the size of the mapping population used to identify the QTLs. The QTLProj command of the software was used to homothetically project the peak positions and confidence intervals of each individual QTL onto the consensus map (Gudi et al., 2022). If a specific study did not provide the CI for a particular QTL, we employed population-specific formulas to calculate the 95% CI as follows:

For the $F_{2:3}$ and Backcross populations, $CI(95\,\%) = 530/(R^2 \times N)$

For the RIL population, $CI\ (95\,\%) = 163/(R^2 \times N)$

For the DH population, $CI\ (95\,\%) = 287/(R^2 \times N)$

Where R^2 is the phenotypic variance explained by the individual QTL.

N is the population size.

Meta-analysis of the QTLs

QTL meta-analysis for individual wheat chromosomes was conducted using BioMercator v4.3.2 (Goffinet and Gerber, 2000; Veyrieras et al., 2007). Two different approaches were used, while conducting the analysis, depending upon the number of QTLs projected on each chromosome. We followed the method proposed by Goffinet and Gerber (2000), when the count of projected QTLs per chromosome was 10 or fewer and we applied the method outlined by Veyrieras et al. (2007) when the number of QTLs per chromosome exceeded 10. In the initial approach, the selection of the best model was determined by examining the model with the lowest Akaike information criterion (AIC) scores. In the second approach, the optimal model was identified from a range of models, which encompassed the Bayesian information criterion (BIC), AIC, corrected AIC, AIC3, and average weight of evidence (AWE). If a model met the criteria of having the lowest score in at least three of the other models, it was considered the most suitable choice.

Candidate gene discovery within the MQTLs and their comparison with GWAS-MTAs

For identifying the genes underlying the MQTL regions, the sequences of the markers flanking the MQTL regions were retrieved from public data repositories such as Grain Genes (https://wheat.pw.usda.gov/GG3) or Cereals DB (https://www.cerealsdb.uk.net/cerealgenomics/CerealsDB/indexNEW.php) databases. To ascertain the physical locations of markers, we conducted BLASTN searches against the Wheat Chinese Spring IWGSC RefSeq v1.0 genome assembly, available on the Ensembl Plants database (http://plants.ensembl.org/index.html). The "JBrowse-WHEAT URGI database" (https://urgi.versailles.inra.fr/jbrowseiwgsc/) was also used to determine the physical locations/genomic co-ordinates of specific SNP markers. The peak positions of the MQTLs were calculated using the following formula proposed by Saini et al. (2022a). Furthermore, some high-confidence MQTLs (hcMQTLs) were chosen and investigated for the identification of available CGs. The following criteria was used to choose these hcMQTLs- (i) involvement of at least 3 initial QTLs, (ii) LOD score ≥ 3 and (iii) PVE value >10. The 'BioMart' tool, accessible in the Ensembl Plants database, was utilized to retrieve gene models located within a 2 Mb genomic region (1 Mb region on either side of the MQTL peak position). From the InterPro database (https://www.ebi.ac.uk/interpro/), functional annotations for the available gene models were

obtained. The genes were further narrowed down based on their Knet scores (Knetminer.com). Additionally, to confirm the effectiveness of the MQTLs, data on PM resistance from 11 GWAS published between 2017 and 2023 were collected and used. The details of these GWAS studies with respect to the population size, type of wheat, platform used for genotyping the mapping panel are given in Supplementary Table S5. To pinpoint the significant SNPs and/or marker-trait associations (MTAs) as reported in these GWAS studies, their physical locations were determined using methods such as BLASTN searches, referencing databases, or consulting the source papers. This process was similar to how the physical locations of MQTLs were determined. The MTAs identified through GWAS within 5 Mb genomic areas nearby a MQTL were considered to be co-located. This is because wheat exhibits an extensive linkage disequilibrium (LD) decay range, which is approximately 5 Mb (Yang et al., 2021).

Identification of known major resistance genes co-localizing with MQTLs

The nucleotide sequences for previously characterized PM genes or sequences of markers linked to these genes were retrieved from databases such as Grain Genes and NCBI (www.ncbi.nlm.nih.gov). Subsequently, BLAST searches were performed using these sequences against the wheat reference genome available in the Ensembl Plants database. After identifying the physical locations of the genes, their positions were compared with the physical coordinates of the MQTLs to determine whether they co-located with the MQTLs.

Results

Distribution of QTLs associated with powdery mildew on wheat genome

QTLs from 34 individual mapping studies (published from 1996 to 2021) were collected and screened for information related to different types of mapping population used, chromosome number, marker positions, LOD score and PVE values. A total of 222 QTLs from PM resistance traits were available for meta-QTL analysis. The detailed information on these QTLs is given in Supplementary Table S1. The number of QTLs present on individual chromosomes varied from a minimum of 3 QTLs on chromosome 6D to a maximum of 21 QTLs on 2B (Figure 1A). Moreover, the distribution QTLs across the three sub-genomes displayed significant differences, with 79 QTLs located on sub-genome A, 86 on sub-genome B, and 57 on sub-genome D. LOD scores for individual QTLs ranged from 2.1 to 82.6 with an average of 11.3. Most QTLs (83.3% of the total) had their LOD score of< 6 (Figure 1B). The phenotypic variation explained by an individual QTL varied from 2.3 to 90% with an average of 16.4%. Approximately, 36.93% of QTLs exhibited a PVE value of< 10% and only a small fraction (6.75%) had a PVE value of > 40%, which suggests the

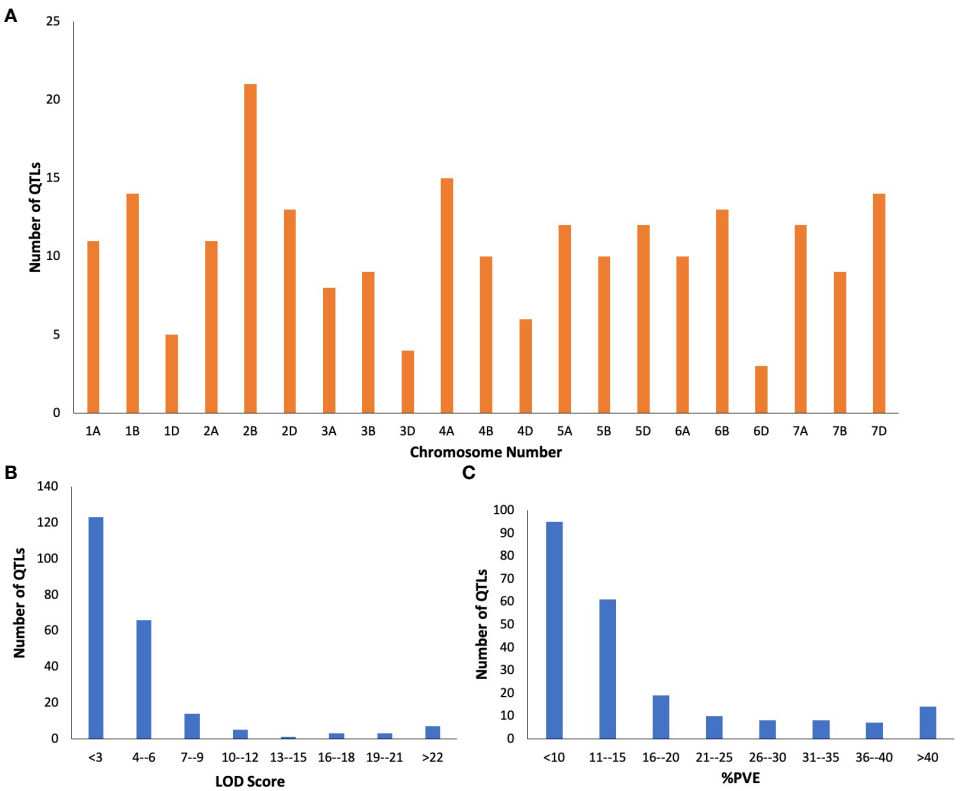


FIGURE 1
Basic characteristics of QTLs associated with PM resistance **(A)** chromosome-wise distribution of QTLs, **(B)** LOD scores of QTLs, **(C)** % PVE of the QTLs.

involvement of both major and minor QTLs governing PM resistance (Figure 1C).

Wheat consensus map and QTL projection

The high-density Wheat_Consensus_Map_2023 generated in the present study using four different genetic maps exhibited a huge variation in the distribution of markers. The marker density on individual chromosomes varied from 1.37 markers/cM on chromosome 4D to 15.92 markers/cM on chromosome 5B, with a mean of 6.40 markers/cM throughout the genome. The marker density on sub-genome B was highest (9.91 markers/cM) followed by sub-genome A (7.21 markers/cM) and sub-genome D (2.54 markers/cM). The length of individual chromosomes also varied significantly (ranged from 157.78 cM for chromosome 4B to 743.48 cM for chromosome 5A with an average of 406.28 cM). The cumulative genetic map length of all the chromosomes in the map was 8531.99 cM spanned by 54,672 markers. On average, there were approximately 2,603 markers mapped per chromosome and the number of different genetic markers that were mapped on a single chromosome ranged from a few hundred (400 on chromosome 4D) to several thousand (4,769 on chromosome 3B) (Figure 2).

Meta-analysis of QTLs associated with powdery mildew resistance

The method proposed by Gerber and Goffinet (Goffinet and Gerber, 2000) was employed for meta-analysis of QTLs across all wheat chromosomes, except for chromosome 1A, 2A, 2B, 2D, 4A, 5D, and 7D, for which the Veyrieras approach was used (Veyrieras et al., 2007), as these had >10 QTLs projected per chromosome. Out of 222 QTLs, only 168 could be projected (75.6% of the total no. of QTLs) onto the newly constructed consensus map. The remaining QTLs could not be projected because of either of the underlying reasons including (i) lack of sufficient number of shared markers between the consensus and initial genetic maps, and (ii) large CI associated with the initial QTLs. A total of 39 MQTLs were predicted for resistance to PM, consisting of 125 initial QTLs out of the total 168 projected QTLs (Supplementary Table S2). The remaining 36 QTLs were singletons (Supplementary Table S3) and therefore they were excluded from subsequent analysis. Further, three QTL hotspots were also identified which consisted of initial QTLs from the same studies. These hotspots were located on chromosomes 2B and 2D (Supplementary Table S4). Across the three sub-genomes, the maximum number of MQTLs were predicted on sub-genome A (18), followed by sub-genome D (12) and sub-genome B (9). Within sub-genome A, chromosome 1A harbored the highest number of MQTLs (4) while chromosome 3A had the lowest number of MQTLs (1). Similarly, for sub-genome B, chromosomes 2B, 3B, 5B, 6B, 7B comprised of 2 MQTLs each while chromosome 1B comprised of only a single MQTL. For sub-genome D, chromosomes 5D and 7D had 4 MQTLs each which is the highest whereas chromosomes 2D and 4D had 2 MQTLs each Figure 3A. No MQTLs were predicted on chromosomes 4B, 1D, 3D and 6D. The number of QTLs per MQTL ranged from ≤ 2 in 16 MQTLs to ≥ 6 QTLs in the 4 MQTLs (viz., MQTL2B.1, MQTL7A.2, MQTL7A.3 and MQTL7D.3) Figure 3B.

The PVE of individual MQTLs varied from a minimum of 5.9 to a maximum of 76.5% with a mean of 21.17% and the LOD score varied from 1.5 to 30.4 with an average of 6.48. Notable features displayed by the initial QTLs, MQTLs and their distribution across

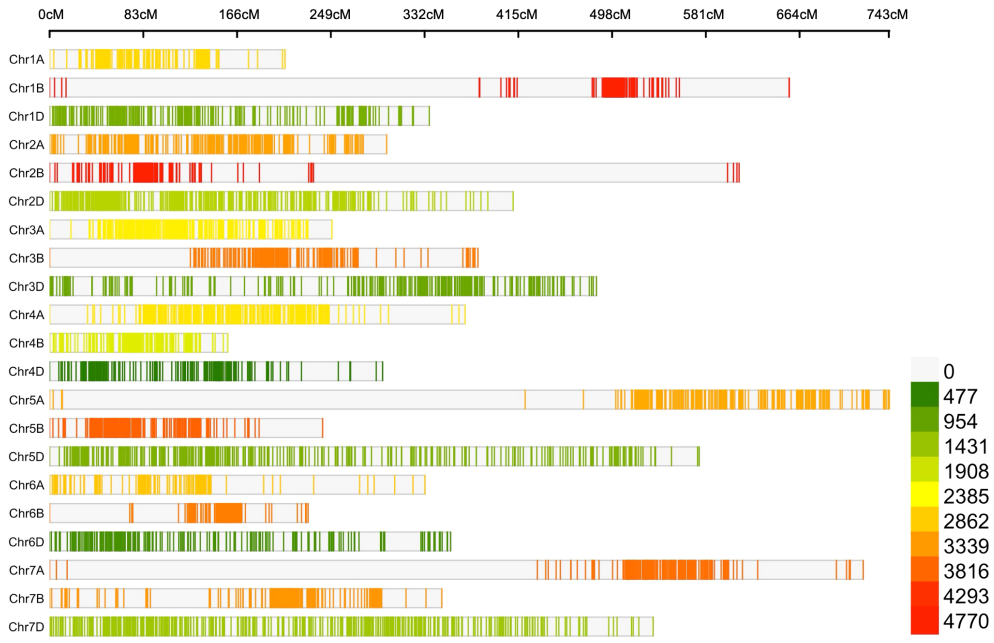


FIGURE 2
Distribution of marker on the consensus map used in MQTL analysis in the current study.

different wheat chromosomes are illustrated in Figures 1, 3, 4, respectively. The CI of the predicted MQTLs and QTL hotspots varied from 0.06 to 28.14 cM and 0.52 to 1.12 respectively. On an average, the CI of MQTLs and QTL hotspots were significantly reduced by a factor of 2.07 and 1.73 respectively compared to the initial QTLs and there were substantial differences in the extent of CI reduction across different wheat chromosomes (Figure 3D). The mean physical CI of the MQTLs was 41.00 Mb, which ranged from 0.000048 Mb (MQTL5B.2) to 380.71 Mb (MQTL6A.2).

Gene models available in MQTL regions

A total of 39 MQTLs were mapped to the physical map of the wheat consensus map used in the present study. However, the exact physical positions of three MQTLs (viz., MQTL2A.3, MQTL7A.2 and MQTL7A.3) could not be ascertained due to the unavailability of nucleotide sequences of the markers flanking these MQTLs (Figure 5, Table 1). To enhance the reliability of the predicted MQTLs, we further refined them, leading to the identification of regions referred to as high-confidence MQTLs (hcMQTLs). In general, each hcMQTL cluster included a minimum of three initial QTLs, with a PVE value greater than 10% and a LOD score greater than three. Further analysis involving gene mining was conducted on 9 hcMQTLs, resulting in the identification of a total of 256 gene models. Among the MQTLs, the one located on chromosome 4D had the maximum number of associated gene models (51). Conversely, the MQTL situated on chromosome 6A had the fewest associated gene models (9) (Supplementary Table S7). These gene models encoded different types of proteins, a few of the important ones are as follows-, including (i) NBS-LRR proteins, (ii) transcription factors (TFs) like MADS box and GRAS, (iii) proteins belonging to oxidoreductase class such as cytochrome P450, (iv) proteins with lectin domain, (v) glycoside hydrolases, and (vi) protein kinases.

Comparison of MQTLs with GWAS-MTAs

The physical coordinates of 39 MQTLs identified in our study were compared to MTAs reported in 10 previous GWAS that comprised a total of 281 MTAs for PM resistance. Among these 39 MQTLs, only 18 were found to be overlapped with GWAS-MTAs (Supplementary Table S5). Some MQTLs were found to be co-localized with MTAs available from multiple GWAS; For instance, MQTL6A.1 co-localized with MTAs identified in three different GWAS. Additionally, three MQTLs (viz., 2B.1, 3B.2 and 5A.2) contained at least 3 initial QTL which were co-localized with multiple MTAs reported from two distinct GWAS.

Co-localization of MQTLs with known PM genes

The identification of the association of known PM genes with individual MQTLs revealed that a total of 2 genes associated with PM resistance, including *Pm2* and *Pm3* were found to be co-localized with 2 MQTLs identified in this study, whereas *Pm8*, *Pm21*, *Pm38* and *Pm41* were found in proximity of MQTLs identified in the present study. (Supplementary Table S6).

Discussion

Multiple QTL mapping studies in wheat related to PM have made significant contributions in advancing our knowledge of the genetic

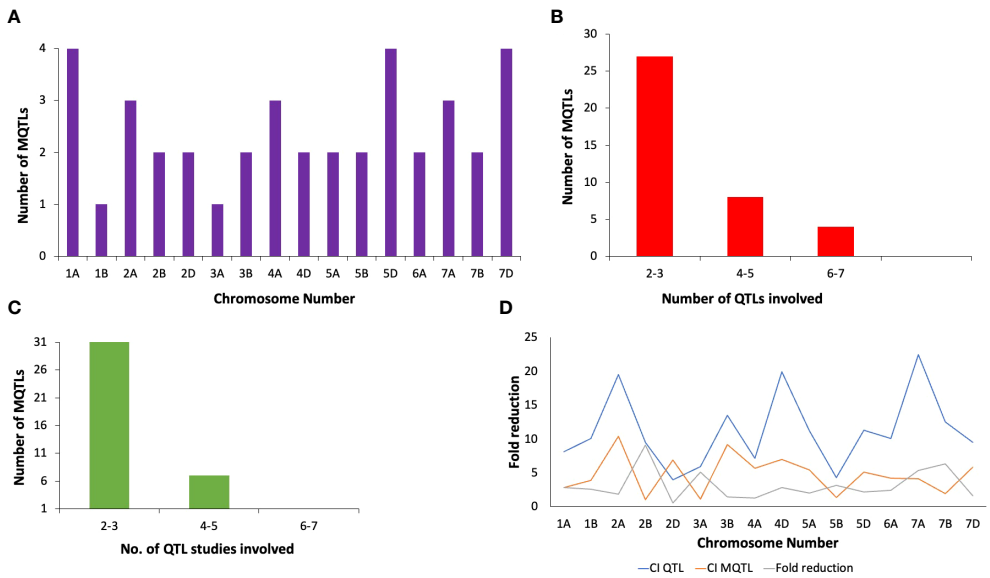


FIGURE 3
Basic characteristics of MQTLs associated with PM resistance **(A)** chromosome wise distribution of MQTLs, **(B)** the number of QTLs involved in different MQTLs, **(C)** the number of QTL studies involved in different MQTLs, **(D)** fold reduction in confidence intervals of QTLs after meta-analysis.

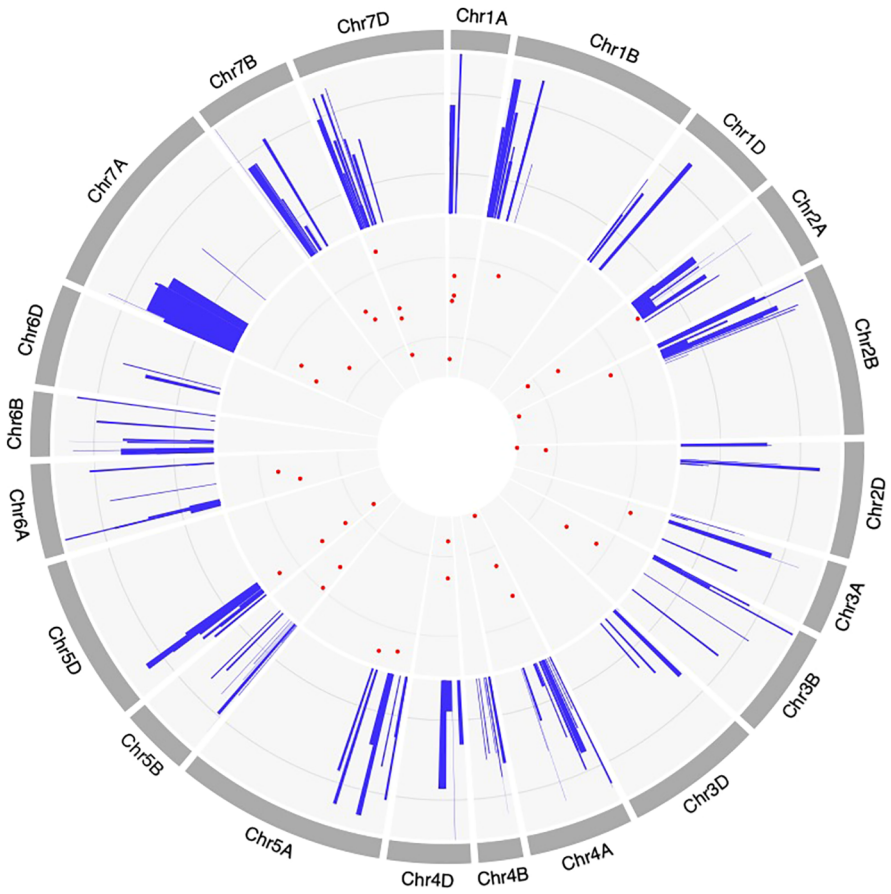


FIGURE 4
Circos diagram representing the features of QTLs and MQTLs associated with powdery mildew resistance. The information projected includes, (moving inwards) the outermost ring represents consensus map, the positions of projected QTLs on the consensus map, and the innermost ring represents the positions of MQTLs on different wheat chromosomes.

basis of quantitative resistance to PM in wheat (Mohler and Stadlmeier, 2019; Xu et al., 2020; Liu et al., 2021). These studies involve the discovery and mapping of distinct genetic regions associated with resistance against PM, offering valuable insights into the genetic complexity of this trait. However, it is widely acknowledged that QTLs identified using one specific mapping population or parental lines may not always perform effectively in a breeding program involving diverse genetic backgrounds (Yang et al., 2021). This phenomenon highlights the need for a more comprehensive and adaptable approach to harness the potential of these QTLs in wheat improvement programme. Overall, MQTL analysis is a powerful method to gather and synthesize information from multiple QTL mapping studies. It provides a more comprehensive understanding of the genetic basis underlying the trait which further helps to refine QTL positions with reduced CI's (Welcker et al., 2011). Meta-QTL analysis has been extensively employed to study a plethora of traits, including disease resistance, across various crop species like rice (Kumar and Nagarajah, 2020; Kumar et al., 2020; Shashiprabha et al., 2022), wheat (Marone et al., 2013; Soriano and Royo, 2015; Venske et al., 2019; Zheng et al., 2021; Pal et al., 2022; Saini et al., 2022b), barley (Schweizer and Stein, 2011) and maize (Rossi et al., 2019). In a previous research effort, MQTLs were identified for wheat's resistance to PM using a limited number of initial QTLs for PM resistance, resulting in the identification of only a small number of MQTLs associated with this trait (Marone et al., 2013). The precision of the meta-analysis findings tends to increase in tandem with the number of initial QTLs employed. It is important to emphasize that a positive relationship typically exists between the precision of results derived from meta-analysis and the number of initial QTLs involved in the analysis (Kumar et al., 2023). Moreover, with the continuous progress in molecular genetics and QTL mapping techniques, there is a consistent discovery and publication of new QTLs. Consequently, it is essential for us to stay updated with these developments to incorporate the latest QTL information into more robust and stable (MQTL) analyses. Hence, in this current study, we conducted a MQTL analysis by integrating QTL data reported in 34 different studies during 1996–2021 on PM resistance and identified 39 MQTLs. to acquire a more profound understanding of how genetic factors regulate resistance to PM in wheat. The first step in the meta-analysis process involved mapping the original QTLs onto a consensus map, which is crucial for identifying common regions of interest through meta-analysis.

10.3389/fpls.2024.1386494

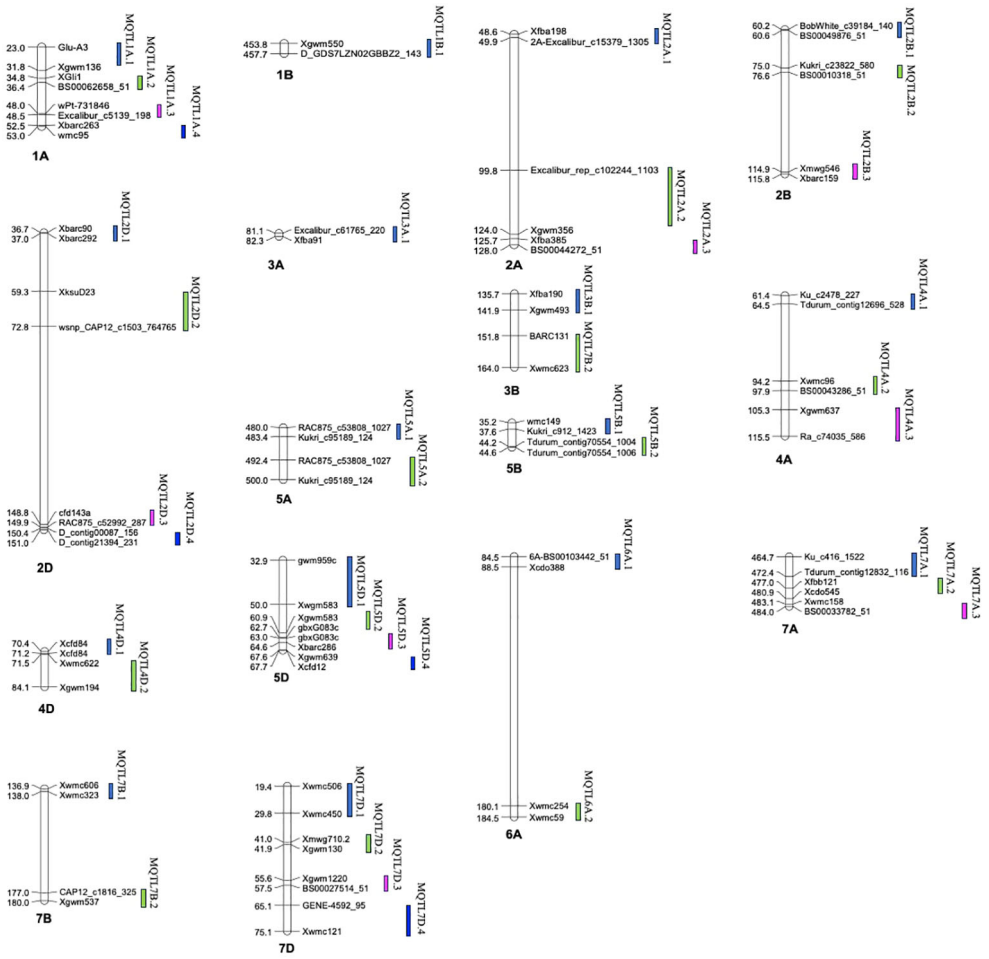


FIGURE 5
Distribution of MQTLs across the different wheat chromosomes.

TABLE 1 MQTLs associated with powdery mildew resistance identified in the present study.

MQTL	Flanking Markers	CI (cM)	LOD Score	PVE%	No. of QTLs involved
MQTL1A.1	*Glu-A3- Xgwm136*	22.98–31.79	5.2	14.4	3
MQTL1A.2	*XGli1- BS00062658_51*	34.78–36.40	10.5	30.2	5
MQTL1A.3	*wPt-731846- Excalibur_c5139_198*	47.98–48.48	2.9	22.9	2
MQTL1A.4	*Xbarc263- wmc95*	52.52–52.95	8.9	46	3
MQTL1B.3	*Xgwm550- D_GDS7LZN02GBBZ2_143*	453.80–457.71	5.8	21.8	4
MQTL2A.1	*Xfba198–2A-Excalibur_c15379_1305*	48.58–49.92	30.4	50.7	4
MQTL2A.2	*Excalibur_rep_c102244_1103- BS00044272_51*	99.84–127.98	12.2	38.2	2
MQTL2A.3	*Xgwm356- Xfba385*	123.95–125.73	8.4	21.9	5
MQTL2B.1	*BobWhite_c39184_140- BS00049876_51*	60.15–60.64	9.1	20.4	6
MQTL2B.2	*Kukri_c23822_580- BS00010318_51*	75.00–76.60	4	9.5	2
MQTL2D.1	*Xbarc90- Xbarc292*	36.69–36.99	3.6	5.9	2
MQTL2D.2	*XksuD23- wsnp_CAP12_c1503_764765*	59.33–72.83	2.7	11.7	2

(Continued)

TABLE 1 Continued

MQTL	Flanking Markers	CI (cM)	LOD Score	PVE%	No. of QTLs involved
MQTL3A.1	*Excalibur_c61765_220- Xfba91*	81.10–82.26	2.8	10.3	2
MQTL3B.1	*Xfba190- Xgwm493*	135.66–141.90	8	15.9	3
MQTL3B.2	*BARC131-Xwmc623*	151.83–164.03	4.6	11.4	3
MQTL4A.1	*Ku_c2478_227- Tdurum_contig12696_528*	61.37–64.45	15.3	22	2
MQTL4A.2	*Xwmc96- BS00043286_51*	94.16–97.93	3.6	36.4	2
MQTL4A.3	*Xgwm637- Ra_c74035_586*	105.34–115.46	3.1	11.8	2
MQTL4D.1	*Xcfd84-Xwmc622*	70.41–71.53	8.9	21.3	2
MQTL4D.2	*Xcfd84- Xgwm194*	71.20–84.14	7.5	18.2	3
MQTL5A.1	*RAC875_c53808_1027- Kukri_c95189_124*	480.00–483.36	3.2	14.2	3
MQTL5A.2	*RAC875_c53808_1027- Kukri_c95189_124*	492.41–499.97	3	18.9	2
MQTL5B.1	*wmc149- Kukri_c912_1423*	35.23–37.58	3.3	8.5	2
MQTL5B.2	*Tdurum_contig70554_1004- Tdurum_contig70554_1006*	44.17–44.56	3.1	10	2
MQTL5D.1	*gwm959c- Xwgm583*	32.86–49.95	3.4	21.2	2
MQTL5D.2	*Xgwm583-gbxG083c*	60.92–62.68	9.4	34.9	5
MQTL5D.3	*gbxG083c- Xbarc286*	63.03–64.55	10.8	76.5	5
MQTL5D.4	*Xgwm639-Xcfd12*	67.63–67.69	2.6	21.7	3
MQTL6A.1	*6A-BS00103442_51-Xcdo388*	84.49–88.52	15.3	35.4	4
MQTL6A.2	*Xwmc254- Xwmc59*	180.11–184.49	8.8	16	2
MQTL7A.1	*Ku_c416_1522- Tdurum_contig12832_116*	464.72–472.43	3.4	13.4	5
MQTL7A.2	*Xfbb121-Xcdo545*	476.97–480.91	3.1	15.2	6
MQTL7A.3	*Xwmc158- BS00033782_51*	483.13–484.00	2.7	14.4	6
MQTL7B.1	*Xwmc606- Xwmc323*	136.90–137.95	4.3	21.3	3
MQTL7B.2	*CAP12_c1816_325- Xgwm537*	177.03–179.95	1.5	14.3	3
MQTL7D.1	*Xwmc506- Xwmc450*	19.39–29.83	5.1	7.1	2
MQTL7D.2	*Xmwg710.2- Xgwm130*	41.02–41.93	3.3	21.1	3
MQTL7D.3	*Xgwm1220- BS00027514_51*	55.57–57.52	5	8.7	6
MQTL7D.4	*GENE-4592_95- Xwmc121*	65.11–75.13	4	12.1	3

Sub-genome B revealed the highest density of genetic markers, thus harboring maximum number of initial QTLs. This finding aligns with previous reports that have explored genetic diversity and the intricate genetic makeup of disease resistance in wheat (Soriano and Royo, 2015; Amo and Soriano, 2022; Pal et al., 2022; Saini et al., 2022b). Conversely, the comparatively low level of genetic variation in the sub-genome D may explain the relatively small number of QTLs detected across various QTL mapping studies which is consistent with previously conducted meta-analyses for disease resistance in wheat that also reported less number of QTLs on sub-genome D (Venske et al., 2019; Liu et al., 2020; Zheng et al., 2021; Pal et al., 2022; Saini et al., 2022a). We believe that our current effort in gathering and analyzing QTL data for PM resistance in wheat represents the most extensive and thorough compilation to date. The highly dense consensus map constructed in our study using four different genetic maps allowed us to identify markers that were closely associated with corresponding MQTLs. The consensus map used in the current study consists of a higher number of markers in comparison to the consensus map utilized in a prior study on MQTL related to PM resistance in wheat, where only 3,618 markers were used and these were obtained by merging only two wheat linkage maps. A higher proportion of QTLs (75.6%) were projected on the consensus map. One potential explanation could be the use of a comprehensive consensus map in the present research. The discovery of 39 MQTLs from the initial pool of 168 QTLs led to a notable decrease by a factor of 2.07 (=10.35/5) in the number of genomic regions or QTLs linked to PM resistance in wheat. Our study stands out as a more up-to-date and comprehensive compilation as compared to a prior meta-analysis on PM resistance

(Marone et al., 2013); for several reasons. Firstly, we incorporated a larger dataset, utilizing 222 QTLs from 34 mapping studies, in contrast to 101 QTLs from 20 studies, which has been shown to improve the accuracy of statistical findings. Secondly, we employed a highly dense consensus map with 54,672 markers, as opposed to the earlier study's use of 3,618 markers. Thirdly, our study integrated a greater number of QTLs (168) into MQTLs due to the use of the dense consensus map. Additionally, we validated 18 MQTLs using GWAS-based MTAs, demonstrating the broader genetic impact of these regions on PM resistance. Moreover, our study employed specific criteria to prioritize hcMQTLs for CG mining, leading to the identification of more promising CGs.

Candidate genes within the hcMQTLs and their association with PM responses

Candidate gene mining within 9 hcMQTLs revealed 256 unique gene models. Twenty-five promising candidate genes were chosen based on Knet score (Table 2). The roles of some genes in conferring resistance to PM is discussed as follows: (i) NBS-LRR domain-containing proteins are also encoded by some cloned *Pm* genes such as *Pm2b*, *Pm60* and *Pm21* that confer PM resistance in wheat (He et al., 2018; Zou et al., 2018; Jin et al., 2022), (ii) Proteins belonging to the protein kinase family are essential components of the defense mechanism in wheat. Receptor-like kinases (RLKs) and plant protection kinases participate in the recognition and initiation of a diverse array of signals connected to various developmental and physiological functions. These include processes related to defense mechanisms as well as beneficial symbiotic interactions (Rentel et al., 2004; Garcia et al., 2012), (iii) Jacalin like lectin domain containing genes have the ability to bind to carbohydrates, recognizing those derived from pathogens or injury during infections (Lannoo and Van Damme, 2014; Esch and Schaffrath, 2017). A mannose-specific JRL (mJRL)-like gene (*TaJRLL1*) which codes for a protein containing two jacalin-like lectin domains has been discovered in wheat. When *TaJRLL1* was introduced into *Arabidopsis thaliana*, resistance towards two fungal pathogens, *F. graminearum* and *Botrytiscinerea* was enhanced. The levels of jasmonic acid (JA) and salicylic acid (SA) showed a substantial increase in transgenic Arabidopsis plants. These results suggest that *TaJRLL1* could be a component of the SA and JA dependent defense signaling pathways (Xiang et al., 2011), (iv) Glycoside hydrolases (GHs) are the most prevalent and widely distributed class of enzymes in fungi, as observed in various studies and their primary function is to enzymatically break down the glycosidic bonds between carbohydrate molecules or between carbohydrates and non-carbohydrate groups (Zhou et al., 2013). One study discovered that GH proteins are recognized by the leucine-rich-repeat receptor-like protein in the dicot plant *Nicotiana benthamiana*. Heterologous expression of this receptor in wheat, made it responsive towards GH proteins, resulting in an increased resistance to *F. graminearum* (a fungal pathogen) and lower levels of the mycotoxin deoxynivalenol in wheat grains (Wang et al., 2023a), (v) Transcription factors like GRAS, MYB and MADS-box play an important role in plant disease resistance. 1R-MYB transcription factor was reported to play key role in disease resistance against stripe rust fungus in wheat (Zhang et al., 2015). Studies by Zhang et al. (2016) and Castelán-Muñoz et al. (2019) have reported that members of the MADS-box gene family participate in the control of both biotic and abiotic stress reactions, which indicates a potential role in stress response. Several MADS-box genes were activated and displayed varying levels of expression after inoculation of wheat spikes with FHB indicating their potential defensive roles against Fusarium infections (Kugler et al., 2013), (vi) The gene *TaCYP72A* which encodes cytochrome P450, plays a significant role enhancing host resistance to Fusarium head blight (FHB) in wheat (Gunupuru et al., 2018), (viii) Synaptobrevin domain containing proteins-TaSYP137 and TaVAMP723, the SNAREs proteins containing streptobrevin and longin domains have been reported to reduce resistance to *Blumeria graminis f.* sp. *Tritici* in wheat (Wang et al., 2023b).

Concordance between MQTLs and known major genes

While there is a record of over 100 PM resistance genes/alleles within 63 loci (*Pm1-Pm66*), only a handful of them have been cloned and thoroughly characterized. These include *Pm1a, Pm2a, Pm2b, Pm3b, Pm5e, Pm8, Pm17, Pm21, Pm24, Pm38, Pm41, Pm46*, and *Pm60*, while the majority remain uncharacterized (Yahiaoui et al., 2004; Krattinger et al., 2009; Moore et al., 2015; Sánchez-Martín et al., 2016; He et al., 2018; Singh et al., 2018; Xing et al., 2018; Zou et al., 2018; Zhang et al., 2019; Li et al., 2020; Lu et al., 2020; Xie et al., 2020; Hewitt et al., 2021; Jin et al., 2022). Except for *Pm24*, *Pm38*, and *Pm46*, which code for a tandem kinase gene known as WTK3, an ABC transporter, and a hexose transporter, respectively, most of the characterized PM resistance genes in wheat are categorized as NLR proteins and exhibit specificity to distinct pathogen races (Mapuranga et al., 2022). As many as 2 PM resistance genes, including *Pm2* and *Pm3* were found to share the same genomic locations as overlap with 2 MQTLs that were identified in this study. For instance, MQTL1A.1 co-localized with *Pm3a* and *Pm3b* and MQTL5D.1 co-localized with both *Pm2a* and *Pm2b* genes conferring resistance to PM in wheat, thus confirming the effectiveness of employing a highly saturated consensus map in MQTL analysis. Furthermore, 4 PM (*Pm8*, *Pm21*, *Pm38* and *Pm41*) genes were found in proximity to the MQTLs identified in the present study.

Recently, Jin et al. (2022) established that the PM resistance genes *Pm2b* was located within the same genomic region as *Pm2a* and *PmCH1357*. However, these genes exhibited distinct resistance profiles, suggesting that *Pm2* exhibits a diverse resistance spectrum among its multiple alleles. The researchers also identified a transcription factor called *TaWRKY76-D* as an interacting partner of *Pm2b*, and this interaction relied on the NB domain of *Pm2b* and the WRKY domain of *TaWRKY76-D*. Interestingly, *TaWRKY76-D* was found to have a negative regulatory effect on PM resistance in wheat. Additionally, the team developed a specific KASP marker known as K529, which offers the advantages of high-throughput

10.3389/fpls.2024.1386494

TABLE 2 Most promising candidate genes associated with powdery mildew resistance.

hcMQTL	Gene Ids	Gene Position (bp)	Knet Score	Functional description
MQTL1A.1	*TraesCS1A03G0022100*	4497200- 4503419	184.25	NB-ARC
MQTL1A.1	*TraesCS1A03G0025200*	5645211- 5650510	39.16	NB-ARC
MQTL1A.1	*TraesCS1A03G0026100*	6007940- 6012646	16.07	NB-ARC
MQTL1A.2	*TraesCS1A03G0027400*	6293918- 6294835	21.76	Leucine-rich repeat domain superfamily
MQTL1A.2	*TraesCS1A03G0032400*	7311700- 7312269	139.08	Transcription factor, MADS-box
MQTL2B.1	*TraesCS2B03G0094000*	22781478- 22785668	32.12	NB-ARC
MQTL2B.1	*TraesCS2B03G0098400*	23575924- 23578715	24.98	Glycoside hydrolase, family 19, catalytic
MQTL2B.1	*TraesCS2B03G0098900*	23596456- 23597365	25.95	Jacalin-like lectin domain
MQTL3B.1	*TraesCS3B03G0631000*	381767017- 381759928	23.74	SANT/Myb domain
MQTL3B.2	*TraesCS3B03G1090700*	681100681- 681097254	151.3	Transcription factor GRAS
MQTL3B.2	*TraesCS3B03G1087500*	679645157- 679642968	21.6	Transcription factor, MADS-box
MQTL4D.2	*TraesCS4D03G0817500*	503968880- 503969956	16.86	BTB/POZ domain
MQTL4D.2	*TraesCS4D03G0815200*	503615478- 503621888	13.89	EF-hand domain
MQTL4D.2	*TraesCS4D03G0814000*	503355740- 503359344	13.89	Palmitoyl protein thioesterase
MQTL5A.1	*TraesCS5A03G0006500*	2408514- 2406435	32.12	Synaptobrevin
MQTL5A.1	*TraesCS5A03G0009300*	3177753- 3175334	47.72	Cytochrome P450
MQTL5A.1	*TraesCS5A03G0009400*	3197691- 3194780	15.6	Cytochrome P450
MQTL5A.1	*TraesCS5A03G0009500*	3218868- 3215100	19.86	3-beta hydroxysteroid dehydrogenase/isomerase
MQTL5A.1	*TraesCS5A03G0009700*	3294536- 3281861	19.86	Terpene synthase, conserved site
MQTL5A.1	*TraesCS5A03G0010300*	3374577- 3370704	33.82	Protein kinase domain
MQTL6A.1	*TraesCS6A03G0385400*	149962969- 149954398	9.1	T-complex 11
MQTL6A.1	*TraesCS6A03G0386400*	150987049- 150982707	10.56	Fumarylacetoacetase
MQTL7D.4	*TraesCS7D03G0436800*	151917055- 151915776	13.89	Plant peroxidase
MQTL7D.4	*TraesCS7D03G0437800*	152121744- 152120457	10.08	Glycoside hydrolase family 16
MQTL7D.4	*TraesCS7D03G0438700*	152949088- 152947488	15.25	NAC domain

and high-efficiency for the detection of *Pm2b*. In the current study, MQTL1A.1 was discovered to co-localize with both *Pm3a* and *Pm3b*, suggesting a connection between these genes and reinforcing the fact that 10 alleles providing race-specific resistance to PM exist at the *Pm3* locus in hexaploid wheat.

Assessing the effectiveness of MQTL through GWAS

GWAS is a promising approach for exploring intricate traits, leveraging both recent and past recombination events within the association panel, thereby enabling precise mapping of these traits (Bush and Moore, 2012; Saini et al., 2022c; Singh et al., 2023; Reddy et al., 2023). The emergence of cost-effective, high-throughput sequencing technologies have simplified the discovery of MTAs associated with various disease resistance traits using genome-wide variants (Kumar et al., 2020; Alemu et al., 2021; Tomar et al., 2021; Nannuru et al., 2022; Vikas et al., 2022; Pradhan et al., 2023). In this study, approximately 46% (18 out of 39) of the identified MQTLs were confirmed through associations with MTAs related to PM resistance. The limited verification of MQTLs by GWAS-MTAs may be attributed to various factors or reasons. Firstly, neither of these methods (MQTL or GWAS) comprehensively captures all the genetic diversity inherent

within the crop species. Secondly, there is a notable disparity in the genetic materials utilized between these two approaches. Furthermore, it is important to note that GWAS primarily targets the identification of common or frequent genetic variants, typically those with a minor allele frequency exceeding 5%. Additionally, environmental factors can significantly influence trait expression, and if not adequately controlled for, they may obscure associations between markers and traits. On the contrary, linkage-based interval mapping studies excel in detecting rare alleles that significantly influence the phenotype.

The stability and reliability of MQTLs could be enhanced when they are corroborated by MTAs identified through multiple GWAS studies and encompass numerous QTLs derived from different interval mapping studies. In our study, four such MQTLs (MQTL2B.1, MQTL3B.2, MQTL6A.1, MQTL5A.2) with at least 2 initials QTLs from different studies with reduced CI (95%) (average CI< 6cM) were verified with multiple MTAs obtained from different GWAS. These meta-QTLs have immense potential and can be regarded as valuable candidates for use in marker-assisted breeding (MAB) initiatives aimed at enhancing PM resistance in wheat.

Conclusions

Our current could identify the most stable and reliable QTLs associated with PM resistance in wheat, providing insight into the intricate quantitative genetic structure of PM resistance. Out of the 39 MQTLs which were identified, 9 hcMQTLs were selected for further investigation to identify the underlying CGs. A total of 256 unique candidate genes were identified within the hcMQTLs, with 25 promising candidates belonging to various gene families known to play roles in disease resistance, such as NBS-LRR, protein kinases, jacalin-like lectins, glycoside hydrolases, and transcription factors. The co-localization of MQTLs with known powdery mildew resistance genes (*Pm2, Pm3, Pm8, Pm21, Pm38*, and *Pm41*) validated the effectiveness of the MQTL analysis. The obtained MQTLs would be useful for further understanding of the molecular mechanisms of PM resistance and for the development of PM-resistant wheat varieties. Additionally, information pertaining to markers flanking the MQTLs can be integrated into genomic selection models, thereby, improving the accuracy of PM resistance through more accurate estimates of genomic estimated breeding values (GEBVs). In the future, breeders have the opportunity to enhance PM resistance trait in wheat by utilizing the most promising MQTLs, specifically 1A.1, 1A.2, 2B.1, 3B.1, 3B.2, 4D.2, 5A.1, 6A.1 and 7D.4 as identified in this research. The candidate gene underlying these MQTLs could bevalidated for marker-assisted breeding programs aimed at enhancing PM resistance in wheat.

Data availability statement

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

Author contributions

DS: Writing – review & editing, Writing – original draft, Software, Data curation, Conceptualization. NB: Writing – review & editing, Software, Conceptualization. AK: Writing – original draft, Software, Data curation. DKS: Writing – review & editing, Supervision, Conceptualization. AS: Writing – original draft, Data curation. AY: Writing – original draft, Data curation. RM: Writing – review & editing, Supervision. AKS: Writing – review & editing, Supervision. VKV: Writing – review & editing, Supervision. GS: Writing – review & editing, Supervision. DK: Writing – review & editing, Supervision, Conceptualization.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. The authors are grateful to the Indian Council of Agricultural Research, Government of India, for supporting this research under the ICAR-NBPGR-DBT Wheat Network Project (No. BT/Ag/Network/Wheat/2019–20).

Acknowledgments

Authors are highly thankful to Director, ICAR-NBPGR for his constant support to DBT-Wheat Network Project. Authors are also thankful to Department of Biotechnology for providing an opportunity to work in this project.

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2024.1386494/full#supplementary-material

References

Alemu, S. K., Huluka, A. B., Tesfaye, K., Haileselassie, T., and Uauy, C. (2021). Genome-wide association mapping identifies yellow rust resistance loci in Ethiopian durum wheat germplasm. *PloS One* 16, e0243675. doi: 10.1371/journal.pone.0243675

Amo, A., and Soriano, J. M. (2022). Unraveling consensus genomic regions conferring leaf rust resistance in wheat *via* meta-QTL analysis. *Plant Genome* 15, e20185. doi: 10.1002/tpg2.20185

Bapela, T., Shimelis, H., Terefe, T., Bourras, S., Sanchez-Martin, J., Douchkov, D., et al. (2023). Breeding wheat for powdery mildew resistance: genetic resources and methodologies—A review. *Agronomy* 13, 1173. doi: 10.3390/agronomy13041173

Bennett, F. G. (1984). Resistance to powdery mildew in wheat: a review of its use in agriculture and breeding programmes. *Plant Pathol. J.* 33, 279–300. doi: 10.1111/j.1365-3059.1984.tb01324.x

Bhullar, N. K., Street, K., Mackay, M., Yahiaoui, N., and Keller, B. (2009). Unlocking wheat genetic resources for the molecular identification of previously undescribed functional alleles at the Pm3 resistance locus. *Proc. Natl. Acad. Sci.* 106, 9519–9524. doi: 10.1073/pnas.0904152106

Bush, W. S., and Moore, J. H. (2012). Chapter 11: Genome-wide association studies. *PloS Comput. Biol.* 8, e1002822. doi: 10.1371/journal.pcbi.1002822

Cao, A., Xing, L., Wang, X., Yang, X., Wang, W., Sun, Y., et al. (2011). Serine/threonine kinase gene Stpk-V, a key member of powdery mildew resistance gene Pm21, confers powdery mildew resistance in wheat. *Proc. Natl. Acad. Sci.* 108, 7727–7732. doi: 10.1073/pnas.1016981108

Castelán-Muñoz, N., Herrera, J., Cajero-Sánchez, W., Arrizubieta, M., Trejo, C., García-Ponce, B., et al. (2019). MADS-box genes are key components of genetic regulatory networks involved in abiotic stress and plastic developmental responses in plants. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.00853

Cheng, P., Guo, M., Hao, X., Guo, X., Yao, Q., Guo, Q., et al. (2022). Evaluation of powdery mildew resistance and molecular detection of resistance genes in an international wheat collection. *Crop Prot.* 160, 106033. doi: 10.1016/j.cropro.2022.106033

Dean, R., Van Kan, J. A., Pretorius, Z. A., Hammond-Kosack, K. E., Di Pietro, A., Spanu, P. D., et al. (2012). The Top 10 fungal pathogens in molecular plant pathology. *Mol. Plant Pathol.* 13, 414–430. doi: 10.1111/j.1364-3703.2011.00783.x

Endelman, J. B., and Plomion, C. (2014). LPmerge: an R package for merging genetic maps by linear programming. *Bioinformatics* 30, 1623–1624. doi: 10.1093/bioinformatics/btu091

Esch, L., and Schaffrath, U. (2017). An update on jacalin-like lectins and their role in plant defense. *Int. J. Mol. Sc* 18, 1592. doi: 10.3390/ijms18071592

Friebe, B., Heun, M., Tuleen, N., Zeller, F. J., and Gill, B. S. (1994). Cytogenetically monitored transfer of powdery mildew resistance from rye into wheat. *Crop Sci.* 34, 621–625. doi: 10.2135/cropsci1994.0011183X003400030003x

Garcia, A. V., Al-Yousif, M., and Hirt, H. (2012). Role of AGC kinases in plant growth and stress responses. *Cell. Mol. Life Sci.* 69, 3259–3267. doi: 10.1007/s00018-012-1093-3

Gaurav, K., Arora, S., Silva, P., Sánchez-Martín, J., Horsnell, R., Gao, L., et al. (2022). Population genomic analysis of Aegilops tauschii identifies targets for bread wheat improvement. *Nat. Biotechnol.* 40, 422–431. doi: 10.1038/s41587-021-01058-4

Goffinet, B., and Gerber, S. (2000). Quantitative trait loci: a meta-analysis. *Genetics* 155, 463–473. doi: 10.1093/genetics/155.1.463

Gunupuru, L. R., Arunachalam, C., Malla, K. B., Kahla, A., Perochon, A., Jia, J., et al. (2018). A wheat cytochrome P450 enhances both resistance to deoxynivalenol and grain yield. *PloS One* 13, e0204992. doi: 10.1371/journal.pone.0204992

He, H., Zhu, S., Zhao, R., Jiang, Z., Ji, Y., Ji, J., et al. (2018). Pm21, encoding a typical CC-NBS-LRR protein, confers broad-spectrum resistance to wheat powdery mildew disease. *Mol. Plant* 11, 879–882. doi: 10.1016/j.molp.2018.03.004

Hewitt, T., Müller, M. C., Molnár, I., Mascher, M., Holušová, K., Šimková, H., et al. (2021). A highly differentiated region of wheat chromosome 7AL encodes a Pm1a immune receptor that recognizes its corresponding AvrPm1a effector from Blumeria graminis. *New Phytol.* 229, 2812–2826. doi: 10.1111/nph.17075

Hurni, S., Brunner, S., Stirnweis, D., Herren, G., Peditto, D., McIntosh, R. A., et al. (2014). The powdery mildew resistance gene Pm8 derived from rye is suppressed by its wheat ortholog Pm3. *Plant J.* 79, 904–913. doi: 10.1111/tpj.12593

Jakobson, I., Reis, D., Tiidema, A., Peusha, H., Timofejeva, L., Valárik, M., et al. (2012). Fine mapping, phenotypic characterization and validation of non-race-specific resistance to powdery mildew in a wheat–*Triticum militinae* introgression line. *Theor. Appl. Genet.* 125, 609–623. doi: 10.1007/s00122-012-1856-0

Jan, I., Saripalli, G., Kumar, K., Kumar, A., Singh, R., Batra, R., et al. (2021). Meta-QTL analysis for stripe rust resistance in wheat. *Sci. Rep*. doi: 10.21203/rs.3.rs-380807/v1

Jin, Y., Liu, H., Gu, T., Xing, L., Han, G., Ma, P., et al. (2022). PM2b, a CC-NBS-LRR protein, interacts with TaWRKY76-D to regulate powdery mildew resistance in common wheat. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.973065

Koller, T., Brunner, S., Herren, G., Hurni, S., and Keller, B.. (2018). Pyramiding of transgenic Pm3 alleles in wheat results in improved powdery mildew resistance in the field. *Theor. Appl. Genet.* 131, 861–871. doi: 10.1007/s00122-017-3043-9

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

Kugler, K. G., Siegwart, G., Nussbaumer, T., Ametz, C., Spannagl, M., Steiner, B., et al. (2013). Quantitative trait loci-dependent analysis of a gene co-expression network associated with Fusarium head blight resistance in bread wheat (*Triticum aestivum* L.). *BMC Genom.* 14, 1–15. doi: 10.1186/1471-2164-14-728

Kumar, D., Kumar, A., Chhokar, V., Gangwar, O. P., Bhardwaj, S. C., Sivasamy, M., et al. (2020). Genome-wide association studies in diverse spring wheat panel for stripe, stem, and leaf rust resistance. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.00748

Kumar, I. S., and Nadarajah, K. (2020). A meta-analysis of quantitative trait loci associated with multiple disease resistance in rice (*Oryza sativa L.*). *Plants* 9, 1491. doi: 10.3390/plants9111491

Kumar, S., Saini, D. K., Jan, F., Jan, S., Tahir, M., Djalovic, I., et al. (2023). Comprehensive meta-QTL analysis for dissecting the genetic architecture of stripe rust resistance in bread wheat. *BMC Genom.* 24, 1–19. doi: 10.1186/s12864-023-09336-y

Kumar, S., Singh, V. P., Saini, D. K., Sharma, H., Saripalli, G., Kumar, S., et al. (2021). Meta-QTLs, ortho-MQTLs, and candidate genes for thermotolerance in wheat (*Triticum aestivum* L.). *Mol. Breed.* 41, 1–22. doi: 10.1007/s11032-021-01264-7

Kumari, A., Sharma, D., Sharma, P., Wang, C., Verma, V., Patil, A., et al. (2023). Meta-QTL and haplo-pheno analysis reveal superior haplotype combinations associated with low grain chalkiness under high temperature in rice. *Front. Plant Sci.* 14. doi: 10.3389/fpls.2023.1133115

Lannoo, N., and Van Damme, E. J. (2014). Lectin domains at the frontiers of plant defense. *Front. Plant Sci.* 5. doi: 10.3389/fpls.2014.00397

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

Liu, S., Hall, M. D., Griffey, C. A., and McKendry, A. L. (2009). Meta-analysis of QTL associated with Fusarium head blight resistance in wheat. *Crop Sci.* 49, 1955–1968. doi: 10.2135/cropsci2009.03.0115

Liu, Y., Salsman, E., Wang, R., Galagedara, N., Zhang, Q., Fiedler, J. D., et al. (2020). Meta-QTL analysis of tan spot resistance in wheat. *Theor. Appl. Genet.* 133, 2363–2375. doi: 10.1007/s00122-020-03604-1

Liu, Z., Wang, Q., Wan, H., Yang, F., Wei, H., Xu, Z., et al. (2021). QTL mapping for adult-plant resistance to powdery mildew in Chinese elite common wheat Chuanmai104. *Cereal Res. Commun.* 49, 99–108. doi: 10.1007/s42976-020-00082-5

Löffler, M., Schön, C. C., and Miedaner, T. (2009). Revealing the genetic architecture of FHB resistance in hexaploid wheat (*Triticum aestivum* L.) by QTL meta-analysis. *Mol. Breed.* 23, 473–488. doi: 10.1007/s11032-008-9250-y

Lu, P., Guo, L., Wang, Z., Li, B., Li, J., Li, Y., et al. (2020). A rare gain of function mutation in a wheat tandem kinase confers resistance to powdery mildew. *Nat. Commun.* 11, 680. doi: 10.1038/s41467-020-14294-0

Maccaferri, M., Zhang, J., Bulli, P., Abate, Z., Chao, S., Cantu, D., et al. (2015). A genome-wide association study of resistance to stripe rust (Puccinia striiformis f. sp. tritici) in a worldwide collection of hexaploid spring wheat (*Triticum aestivum* L.). *G3: Genes Genomes Genet.* 5, 449–465. doi: 10.1534/g3.114.014563

Mapuranga, J., Chang, J., and Yang, W. (2022). Combating powdery mildew: Advances in molecular interactions between Blumeria graminis f. sp. tritici and wheat. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.1102908

Marone, D., Russo, M. A., Laidò, G., De Vita, P., Papa, R., Blanco, A., et al. (2013). Genetic basis of qualitative and quantitative resistance to powdery mildew in wheat: from consensus regions to candidate genes. *BMC Genom.* 14, 1–17. doi: 10.1186/1471-2164-14-562

Maurya, S., Singh, M., Kumar, S., Lakhran, L., Kumar, N., and Prajapati, S. (2021). Breeding approaches for disease resistance in crop plants: A review. *Ann. Clin. Lab. Sci.* 4, 1022.

Mohler, V., and Stadlmeier, M. (2019). Dynamic QTL for adult plant resistance to powdery mildew in common wheat (*Triticum aestivum* L.). *J. Appl. Genet.* 60, 291–300. doi: 10.1007/s13353-019-00518-7

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. doi: 10.1038/ng.3439

Morgounov, A., Tufan, H. A., Sharma, R., Akin, B., Bagci, A., Braun, H. J., et al. (2012). Global incidence of wheat rusts and powdery mildew during 1969–2010 and durability of resistance of winter wheat variety Bezostaya 1. *Eur. J. Plant Pathol.* 132, 323–340. doi: 10.1007/s10658-011-9879-y

Nannuru, V. K. R., Windju, S. S., Belova, T., Dieseth, J. A., Alsheikh, M., Dong, Y., et al. (2022). Genetic architecture of fusarium head blight disease resistance and associated traits in Nordic spring wheat. *Theor. Appl. Genet.* 135, 2247–2263. doi: 10.1007/s00122-022-04109-9

Pal, N., Jan, I., Saini, D. K., Kumar, K., Kumar, A., Sharma, P. K., et al. (2022). Meta-QTLs for multiple disease resistance involving three rusts in common wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* 135, 2385–2405. doi: 10.1007/s00122-022-04119-7

Pal, N., Saini, D. K., and Kumar, S. (2021). Meta-QTLs, ortho-MQTLs and candidate genes for the traits contributing to salinity stress tolerance in common wheat (*Triticum aestivum* L.). *Physiol. Mol. Biol. Plants* 27, 2767–2786. doi: 10.1007/s12298-021-01112-0

Pradhan, A. K., Budhlakoti, N., Chandra Mishra, D., Prasad, P., Bhardwaj, S. C., Sareen, S., et al. (2023). Identification of novel QTLs/defense genes in spring wheat germplasm panel for seedling and adult plant resistance to stem rust and their validation through KASP marker assays. *Plant Dis.* 107 (6), 1847–1860. doi: 10.1094/PDIS-09-22-2242-RE

Reddy, S. S., Saini, D. K., Singh, G. M., Sharma, S., Mishra, V. K., and Joshi, A. K. (2023). Genome-wide association mapping of genomic regions associated with drought stress tolerance at seedling and reproductive stages in bread wheat. *Front. Plant Sci.* 14. doi: 10.3389/fpls.2023.1166439

Rentel, M. C., Lecourieux, D., Ouaked, F., Usher, S. L., Petersen, L., Okamoto, H., et al. (2004). OXI1 kinase is necessary for oxidative burst-mediated signalling in. *Arabidopsis Nat.* 427, 858–861. doi: 10.1038/nature02353

Rossi, E. A., Ruiz, M., Rueda Calderón, M. A., Bruno, C. I., Bonamico, N. C., and Balzarini, M. G. (2019). Meta-analysis of QTL studies for resistance to fungi and viruses in maize. *Crop Sci.* 59, 125–139. doi: 10.2135/cropsci2018.05.0330

Saini, D. K., Chahal, A., Pal, N., Srivastava, P., and Gupta, P. K. (2022a). Meta-analysis reveals consensus genomic regions associated with multiple disease resistance in wheat (*Triticum aestivum* L.). *Mol. Breed.* 42 (3), 11. doi: 10.1007/s11032-022-01282-z

Saini, D. K., Chopra, Y., Singh, J., Sandhu, K. S., Kumar, A., Bazzer, S., et al. (2022c). Comprehensive evaluation of mapping complex traits in wheat using genome-wide association studies. *Mol. Breed.* 42, 1–52. doi: 10.1007/s11032-021-01272-7

Saini, D. K., Srivastava, P., Pal, N., and Gupta, P. K. (2022b). Meta-QTLs, ortho-meta-QTLs and candidate genes for grain yield and associated traits in wheat (*Triticum aestivum L.*). *Theor. Appl. Genet.* 135 (3), 1049–1081. doi: 10.1007/s00122-021-04018-3

Sánchez-Martín, J., Steuernagel, B., Ghosh, S., Herren, G., Hurni, S., Adamski, N., et al. (2016). Rapid gene isolation in barley and wheat by mutant chromosome sequencing. *Genome Biol.* 17, 1–7. doi: 10.1186/s13059-016-1082-1

Sánchez-Martín, J., Widrig, V., Herren, G., Wicker, T., Zbinden, H., Gronnier, J., et al. (2021). Wheat Pm4 resistance to powdery mildew is controlled by alternative splice variants encoding chimeric proteins. *Nat. Plants* 7, 327–341. doi: 10.1038/s41477-021-00869-2

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

Schweizer, P., and Stein, N. (2011). Large-scale data integration reveals colocalization of gene functional groups with meta-QTL for multiple disease resistance in barley. *Mol. Plant-Microbe Interact.* 24, 1492–1501. doi: 10.1094/MPMI-05-11-0107

Sharma, D., Kumari, A., Sharma, P., Singh, A., Sharma, A., Mir, Z. A., et al. (2023). Meta-QTL analysis in wheat: progress, challenges and opportunities. *Theor. Appl. Genet.* 136, 1–25. doi: 10.1007/s00122-023-04490-z

Shashiprabha, H. E. D., Abeysundara, S. P., and Ariyarathna, H. A. C. K. (2022). Meta-QTL analysis identified stable quantitative trait loci (QTLs) and associated resistance gene analogues in rice. *J. Natl. Sci. Found* 50, 35–51. doi: 10.4038/jnsfsr.v50i1.10372

Simeone, R., Piarulli, L., Nigro, D., Signorile, M. A., Blanco, E., Mangini, G., et al. (2020). Mapping powdery mildew (Blumeria graminis f. sp. tritici) resistance in wild and cultivated tetraploid wheats. *Int. J. Mol. Sci.* 21 7910. doi: 10.3390/ijms21217910

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

Singh, K., Saini, D. K., Saripalli, G., Batra, R., Gautam, T., Singh, R., et al. (2022). WheatQTLdb V2. 0: A supplement to the database for wheat QTL. *Mol. Breed.* 42, 56. doi: 10.1007/s11032-022-01329-1

Singh, G., Saini, D. K., Singh, G., Halladakeri, P., Kumar, P., Shamshad, M., et al. (2022). Unravelling consensus genomic regions associated with quality traits in wheat using meta-analysis of quantitative trait loci. *Planta* 255, 115. doi: 10.1007/s00425-022-03904-4

Singh, G., Saini, D. K., Halladakeri, P., Singh, G., Singh, S., Kaur, S., et al. (2023). Genome-wide association study unravels genomic regions associated with chlorophyll fluorescence parameters in wheat (*Triticum aestivum* L.) under different sowing conditions. *Plant Cell Rep.* 42, 1453–1472. doi: 10.1007/s00299-023-03041-6

Somers, D. J., Isaac, P., and Edwards, K. (2004). A high-density microsatellite consensus map for bread wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* 109, 1105–1114. doi: 10.1007/s00122-004-1740-7

Song, Q. J., Shi, J. R., Singh, S., Fickus, E. W., Costa, J. M., Lewis, J., et al. (2005). Development and mapping of microsatellite (SSR) markers in wheat. *Theor. Appl. Genet.* 110, 550–560. doi: 10.1007/s00122-004-1871-x

Soriano, J. M., and Royo, C. (2015). Dissecting the genetic architecture of leaf rust resistance in wheat by QTL meta-analysis. *Phytopathology* 105, 1585–1593. doi: 10.1094/PHYTO-05-15-0130-R

Sosnowski, O., Charcosset, A., and Joets, J. (2012). BioMercator V3: an upgrade of genetic map compilation and quantitative trait loci meta-analysis algorithms. *Bioinformatics* 28, 2082–2083. doi: 10.1093/bioinformatics/bts313

Srichumpa, P., Brunner, S., Keller, B., and Yahiaoui, N. (2005). Allelic series of four powdery mildew resistance genes at the Pm3 locus in hexaploid bread wheat. *Plant Physiol.* 139, 885–895. doi: 10.1104/pp.105.062406

Tomar, V., Singh, D., Dhillon, G. S., Singh, R. P., Poland, J., Joshi, A. K., et al. (2021). New QTLs for spot blotch disease resistance in wheat (*Triticum aestivum* L.) using genome-wide association mapping. *Front. Genet.* 11. doi: 10.3389/fgene.2020.613217

Venske, E., Dos Santos, R. S., Farias, D. D. R., Rother, V., Da Maia, L. C., Pegoraro, C., et al. (2019). Meta-analysis of the QTLome of Fusarium head blight resistance in bread wheat: refining the current puzzle. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.00727

Veyrieras, J. B., Goffinet, B., and Charcosset, A. (2007). MetaQTL: a package of new computational methods for the meta-analysis of QTL mapping experiments. *BMC Bioinform.* 8, 1–16. doi: 10.1186/1471-2105-8-49

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

Wang, Z. L., Li, L. H., He, Z. H., Duan, X. Y., Zhou, Y. L., Chen, X. M., et al. (2005). Seedling and adult plant resistance to powdery mildew in Chinese bread wheat cultivars and lines. *Plant Dis.* 89, 457–463. doi: 10.1094/PD-89-0457

Wang, Z., Yang, B., Zheng, W., Wang, L., Cai, X., Yang, J., et al. (2023a). Recognition of glycoside hydrolase 12 proteins by the immune receptor RXEG1 confers Fusarium head blight resistance in wheat. *Plant Biotechnol. J.* 21 (4), 769–781. doi: 10.1111/pbi.13995

Wang, G., Zhang, X., Guo, H., Zhao, C., Zhang, H., Chen, C., et al. (2023b). TaSYP137 and TaVAMP723, the SNAREs Proteins from Wheat, Reduce Resistance to Blumeria graminis f. sp. tritici. *Int. J. Mol. Sci.* 24 (5), 4830. doi: 10.3390/ijms24054830

Welcker, C., Sadok, W., Dignat, G., Renault, M., Salvi, S., Charcosset, A., et al. (2011). A common genetic determinism for sensitivities to soil water deficit and evaporative demand: meta-analysis of quantitative trait loci and introgression lines of maize. *Plant Physiol.* 157, 718–729. doi: 10.1104/pp.111.176479

Wen, W., He, Z., Gao, F., Liu, J., Jin, H., Zhai, S., et al. (2017). A high-density consensus map of common wheat integrating four mapping populations scanned by the 90K SNP array. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.01389

Xiang, Y., Song, M., Wei, Z., Tong, J., Zhang, L., Xiao, L., et al. (2011). A jacalin-related lectin-like gene in wheat is a component of the plant defence system. *J. Exp. Bot.* 62, 5471–5483. doi: 10.1093/jxb/err226

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

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

Xu, X., Zhu, Z., Jia, A., Wang, F., Wang, J., Zhang, Y., et al. (2020). Mapping of QTL for partial resistance to powdery mildew in two Chinese common wheat cultivars. *Euphytica* 216, 1–12. doi: 10.1007/s10681-019-2537-8

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

Yahiaoui, N., Kaur, N., and Keller, B. (2009). Independent evolution of functional Pm3 resistance genes in wild tetraploid wheat and domesticated bread wheat. *The Plant Journal.* 57(5), 846–856. doi: 10.1111/j.1365-313X.2008.03731.x

Yang, Y., Amo, A., Wei, D., Chai, Y., Zheng, J., Qiao, P., et al. (2021). Large-scale integration of meta-QTL and genome-wide association study discovers the genomic regions and candidate genes for yield and yield-related traits in bread wheat. *Theor. Appl. Genet.* 134, 3083–3109. doi: 10.1007/s00122-021-03881-4

Yu, L. X., Barbier, H., Rouse, M. N., Singh, S., Singh, R. P., Bhavani, S., et al. (2014). A consensus map for Ug99 stem rust resistance loci in wheat. *Theor. Appl. Genet.* 127, 1561–1581. doi: 10.1007/s00122-014-2326-7

Zhang, Z., Chen, J., Su, Y., Liu, H., Chen, Y., Luo, P., et al. (2015). TaLHY, a 1R-MYB transcription factor, plays an important role in disease resistance against stripe rust fungus and ear heading in wheat. *PloS One* 10, e0127723. doi: 10.1371/journal.pone.0127723

Zhang, Z., Li, H., Qin, G., He, C., Li, B., and Tian, S. (2016). The MADS-Box transcription factor Bcmads1 is required for growth, sclerotia production and pathogenicity of Botrytis cinerea. *Sci. Rep.* 6, 33901. doi: 10.1038/srep33901

Zhang, D., Zhu, K., Dong, L., Liang, Y., Li, G., Fang, T., et al. (2019). Wheat powdery mildew resistance gene Pm64 derived from wild emmer (*Triticum turgidum* var. *dicoccoides*) is tightly linked in repulsion with stripe rust resistance gene Yr5. *Crop J.* 7, 761–770. doi: 10.1016/j.cj.2019.03.003

Zheng, T., Hua, C., Li, L., Sun, Z., Yuan, M., Bai, G., et al. (2021). Integration of meta-QTL discovery with omics: Towards a molecular breeding platform for improving wheat resistance to Fusarium head blight. *Crop J.* 9, 739–749. doi: 10.1016/j.cj.2020.10.006

Zhou, P., Chen, Z., Yan, Q., Yang, S., Hilgenfeld, R., and Jiang, Z. (2013). The structure of a glycoside hydrolase family 81 endo-β-1, 3-glucanase. *Acta Crystallographica Section D: Biol. Crystallogr.* 69, 2027–2038. doi: 10.1107/S090744491301799X

Zou, S., Wang, H., Li, Y., Kong, Z., and Tang, D. (2018). The NB-LRR gene Pm60 confers powdery mildew resistance in wheat. *New Phytol.* 218, 298–309. doi: 10.1111/nph.14964

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 09 July 2024
DOI 10.3389/fpls.2024.1383135

Check for updates

OPEN ACCESS

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

REVIEWED BY
Xuming Li,
Hugo Biotechnologies Co., Ltd., China
Diaga Diouf,
Cheikh Anta Diop University, Senegal

*CORRESPONDENCE
Saad Serfraz
serfraz.saad@gmail.com

RECEIVED 06 February 2024
ACCEPTED 12 June 2024
PUBLISHED 09 July 2024

CITATION
Sultan AB, Nawaz H, Saleem F, Nawaz S, Danial M, Iftikhar R, Maqsood U, Areej A, Shakoor S, Aljarba NH, Maqbool R, Rizwan M and Serfraz S (2024) Divergent evolution of *NLR* genes in the genus *Glycine*: impacts of annuals and perennials' life history strategies. *Front. Plant Sci.* 15:1383135. doi: 10.3389/fpls.2024.1383135

COPYRIGHT
© 2024 Sultan, Nawaz, Saleem, Nawaz, Danial, Iftikhar, Maqsood, Areej, Shakoor, Aljarba, Maqbool, Rizwan and Serfraz. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Divergent evolution of *NLR* genes in the genus *Glycine*: impacts of annuals and perennials' life history strategies

Abu Bakar Sultan[1], Humera Nawaz[2], Fozia Saleem[1], Sehar Nawaz[1], Muhammad Danial[1], Romana Iftikhar[1], Umer Maqsood[3], Amna Areej[1], Sidra Shakoor[1], Nada H. Aljarba[4], Rizwan Maqbool[5], Muhammad Rizwan[1] and Saad Serfraz[1]*

[1]Evolutionary Biology Lab, CABB, University of Agriculture Faisalabad, Faisalabad, Pakistan, [2]Department of Botany, Division of Science and Technology, University of Education, Lahore, Pakistan, [3]Agricultural Biotechnology Division, National Institute for Biotechnology and Genetic Engineering, Faisalabad, Pakistan, [4]Department of Biology, College of Science, Princess Nourah Bint Abdulrahman University, Riyadh, Saudi Arabia, [5]Department of Agronomy, University of Agriculture, Faisalabad, Pakistan

Within the family *Fabaceae*, the genus *Glycine* is composed of two subgenera annuals (2n=40) and perennials. This life strategy transition may have differentially affected the evolution of various gene families. Its cultivated species *G. max* has high level of susceptibility to major pathogens including viruses, bacteria and fungi. Understanding nucleotide-binding domain leucine-rich repeat (*NLR*) genes evolution in soybean is of paramount importance due to their central role in plant immunity and their potential in improving disease resistance in soybean cultivars. In this study, we investigated the significance of this annual-perennial transition on the macroevolution of *NLR* genes in the genus *Glycine*. Our results reveal a remarkable distinction between annual species such as *Glycine max* and *Glycine soja*, which exhibit an expanded *NLR*ome compared to perennial species (*G. cyrtoloba, G. stenophita, G. dolichocarpa, G. falcata, G. syndetika, G. latifolia* and *G. tomentella*). Our evolutionary timescale analysis pinpoints recent accelerated gene duplication events for this expansion, which occurred between 0.1 and 0.5 million years ago, driven predominantly by lineage-specific and terminal duplications. In contrast, perennials initially experienced significant contraction during the diploidisation phase following the *Glycine*-specific whole-genome duplication event (~10 million years ago). Despite the reduction in the *NLR*ome, perennial lineages exhibit a unique and highly diversified repertoire of *NLR* genes with limited interspecies synteny. The investigation of gene gain and loss ratios revealed that this diversification resulted from the birth of novel genes following individual speciation events. Among perennials, *G. latifolia*, a well-known resistance resource, has the highest ratio of these novel genes in the tertiary gene pool. Our study suggests evolutionary mechanisms, including recombination and transposition, as potential drivers for the emergence of these novel genes. This study also provides evidence for the unbalanced expansion of the *NLR*ome in the D^t subgenome compared with the A^t subgenome in the young allopolyploid *G. dolichocarpa*. To the best of our knowledge, this is the first study to investigate

the effect of annuality and perenniality life transition on the evolution of *NLR* genes in the genus *Glycine* to identify its genomics resources for improving the resistance of soybean crop with global importance on the economy and food security.

KEYWORDS

***glycine*, perennials, annuals, *NLR* genes, diversification, evolution**

Introduction

Soya (*Glycine max*), a legume, is the fourth most important agricultural commodity in the world. It is widely cultivated and used to produce biodiesel, high-quality livestock feed and the cheapest edible protein for humans (Sedivy et al., 2017). The genus *Glycine* comprises two subgenera (*Glycine* subgenus and *Soja* subgenus) that diverged approximately 5 million years ago (Mya) and exhibit different life strategies. The subgenus *Soja* consists of annual plants such as the soybean (*Glycine max*) and its wild ancestor, *Glycine soja*, which was domesticated about 6,000–9,000 years ago. These plants are indigenous to eastern Asia, including China, Japan, Korea, and some parts of Russia. In contrast, the subgenus *Glycine* encompasses around 30 perennial species. The majority of these species are found in Australia and inhabit a variety of environments, including deserts, sandy beaches, rocky outcrops, as well as monsoonal, temperate, and subtropical forests (Innes et al., 2008). All diploid *Glycine* species have a chromosome count of (2n = 4x = 38 or 40), which is different from most species in the legume tribe *Phaseoleae*, which usually have (2n = 20 or 22). Both *Glycine* subgenera basically have (2n = 40). More recently, probably within the past few tens of thousands of years, allopolyploid taxa have appeared within the subgenus *Glycine*. These freshly developed allopolyploids of the *Glycine* subgenus have been discovered as far north as Taiwan and the Ryukyu Islands (Egan and Doyle, 2010; Forrester and Ashman 2018). The progenitor of the entire *Glycine* genus underwent a polyploidy event approximately 5–13 million years ago. Within the last 500,000 years, an additional wave of genome duplication has occurred in the subgenus *Glycine*, giving rise to a vast and fully characterized polyploid complex (Sherman-Broyles et al., 2014; Kim et al., 2015). This complex has eight allopolyploid species (2n = 78, 80) and one autopolyploid species (2n = 80), which were produced from diverse combinations of diploid (2n = 38, 40) genomes. Among them, *G. dolichocarpa* (4n=80) is a tetraploid that consists of two subgenomes A^t and D^t, which arose from two progenitor species *G. syndetika* (A) and *G. tomentella* (D) (Li et al., 2014; Manzoor et al., 2019).

A reference genome for the cultivated soybean (*G. max*) accession 'Williams 82' was released in 2010 (Schmutz et al., 2010), while reference genomes for its annual relative G. *soja* accession IT182932 and W05 were published in 2019 (Xie et al., 2019). The genomes of *Glycine* species were divided into seven categories based on their potential to form fertile hybrids and the degree of pairing of meiotic chromosomes. These groups were labeled with capital letters A through G (Sherman-Broyles et al., 2014; Kim et al ., 2015). The annual wild cousin of soybean (*Glycine soja*) and the cultivated soybean (*Glycine max*) both belong to the G genome group and have the same number of chromosomes (2n = 40).

The perennial wild relatives of soybeans possess significantly more complex genomes compared to their annual counterparts, which are categorized into six distinct genome groups (A–F). Members of the perennials, subgenus *Glycine* have 2n = 38, 40, 78, and 80 chromosomes, in contrast to the subgenus *Soja* (Sherman-Broyles et al., 2014). Perennial soybeans have not had as much genomic study as their annual counterparts, despite efforts to explore their genetic properties.

Among plant resistance genes, nucleotide-binding domain leucine-rich repeat (*NLR*) genes are the principal component of the effector-triggered immunity (ETI) (Cui et al 2015). Through these intracellular *NLR* genes, resistance reactions result in infected plant tissues, typically accompanied by hypersensitivity reactions (HRs) (Pan et al., 2000; Jones et al., 2016). When invading pathogens are detected, these *NLR* proteins become activated, leading to changes in the conformation of the nucleotide-binding site (NBS) domain. To stop infected cells from spreading, the N-terminal domains of the exposed NBS domain start a downstream hypersensitive response (HR) that causes apoptosis to stop the proliferation and transmission of the pathogen (Andersen et al., 2018; Wang and Chai 2020). Based on the type of N-terminal domain, *NLR* genes are divided into four subclasses: Toll/Interleukin-1 receptor-like (TIR-*NLR*), RX-type coiled-coil (CC-*NLR*), CC_R-*NLR* subclade with RTP8-type CC domain (CC_R-*NLR*), and G10 subclade (CC_{G10}-*NLR*), a recently proposed category with a distinct type of CC that forms a monophyletic group. Given the global dominance of soybeans, there is a growing need to enhance their genetic potential. Both domesticated (*G. max*) and wild (*G. soja*) annual species show lower levels of genetic diversity (Quershi et al ., 2023).

Globally soybean production is significantly hampered due to high susceptibility of *G. max* to viral, bacterial, fungal and nematode pathogens. In the US alone, an estimated loss of $95.48 billion were

occurred during 1996 to 2016 highlighting the significant financial implication of these pathogens on soybean production (Bandara et al., 2020). Given the economic importance of soybean and the significant yield losses caused by various diseases, understanding the NLR genes in soybean is vital for crop improvement programs (Araújo et al., 2019). It enables the development of soybean varieties with enhanced disease resistance through traditional breeding or biotechnological approaches. For example, the identification and functional characterization of NLR genes can lead to the development of soybean lines harboring single or multiple NLR-encoding R gene receptors, potentially offering broad-spectrum resistance (Araújo et al., 2019). The perennial wild relatives of the subgenus *Glycine* are a valuable resource due to their disease resistance genes and adaptability to various habitats. To identify the unique mechanism of *NLR* gene evolution, it is necessary to perform a comprehensive characterization of *NLR* genes in both annual and perennial species of *Glycine*. The major objectives of this study to investigate the long-term evolutionary history of NLR genes in annuals and perennials species, identification of novel resistance resources in genus *Glycine* and effect of diploidization on their evolution. This study provides insights into the complexity and evolution of *NLR* genes in *Glycine* species, which could enhance disease resistance in *Glycine* crops and aid in the development of more resilient soybean cultivars.

Materials and methods

Mining of *NLR* genes in for *Glycine* species

The genome sequences of various species within the *Glycine* genus can be accessed through several genome database portals. For this study, we prioritized chromosomal anchored genomes and annotation files for all nine *Glycine* species and acquired them from two portals (NCBI and legume-info) database (Supplementary Table S1). We downloaded the genomes of *G. max, G. cyrtoloba, G. stenophita, G. dolichocarpa, G. falcata, G. syndetika, G. latifolia, G. tomentella, and G. soja* (Supplementary Table S1). To enable comparison of ancestral species, G. *dolichocarpa*, a tetraploid species, was divided into two distinct subgenomes, A^t and D^t. All genome files were labelled into their respective transcriptome, proteome, and gene transfer file formats. The reference proteomes of all nine *Glycine* genomes were processed using the *NLR*tracker pipeline (Jones et al., 2016; Kourelis et al., 2021). The *NLR*tracker produces output files containing sequences of *NLR* and *NLR*-associated sequences, as well as annotations of *NLR*, NBARC, deduplicated NBARC sequences, and domain sequences. These files were generated using interproscan and specified motif patterns. To ensure clarity, each NLR gene was subjected to manual curation using clustering and phylogenetic analysis, given the unclear nature of NLRtracker's CCR-NLR annotation.

Phylogeny and classification of *NLR* Genes

A library of NBARC domains has been produced by the PRG database, which includes reference *NLR* genes (Calle García et al., 2022). As previously mentioned, domain clusters were created using UCLUST's 70% identity criteria (Edgar, 2010). The group category was assigned to each cluster using the subgroup nomenclature by (Seo et al., 2016). Seo et al. (2016) used classified clusters as seed probes to extract the NBARC domains of the *Glycine* genus produced by *NLR*tracker. These domains were then aligned with the NBARC seed probes to facilitate a comprehensive phylogenetic study of *Glycine*. Insights into evolution were obtained using the maximum likelihood technique of IQtree v 2.0 (Nguyen et al., 2015). The best fit model, VT + F + R10, was chosen along with 1000 bootstrap repetitions as the adjusted value.

Loci maps and syntenic maps of *NLR*-genes

The Interproscan tool was used to generate various output files, such as the GFF3 file, NLR fasta file, and an assembled file for gene density maps. After that, we used an adjusted bin size of 5 kb to intersect NLR gene sequences in the annotated file using BEDtools (Quinlan and Hall, 2010). After completing this procedure, count files were produced, which were subsequently modified by assigning each coordinate a bin number. Using the Rldeogram package, this generated a karyotype file for display in R (Hao et al., 2020).

Evolutionary analysis in *Glycine NLR*s

As previously stated, Clustalw was used to align the deduced protein sequences of paralogs with their compatible subgroups (Li, 2003). The alignment of the respective nucleotide sequences was performed using the Perl-based pal2nal software (Suyama et al., 2006). For improved alignment, gaps and codons were removed, and Ks values were determined using the MA method and Kaks calculator. Depending on how nucleotide and protein sequences are substituted, Kaks values can be either non-synonymous (changing over time) or synonymous (changing over time).The frequency of evolution in various species may be inferred from Ks values. To prevent substitution saturation, Ks values larger than two were eliminated. *NLR* genes were grouped using the Orthovenn2 program to investigate orthologs. Orthovenn2 discovered common genes across all species by supplying the protein sequences of suspected *NLR* genes of various species with an E-value of 1e-2 as the default option (Xu et al., 2019). *NLR* genes were submitted to Orthofinder for a thorough examination of orthology (Emms and Kelly, 2019). Tree building was performed using the obtained orthologs. The number of gene gain and loss output was obtained sequentially by CAFE5 using orthogroup and ultrametric tree files as input. The available literature provides a thorough overview of evolutionary analysis (Areej et al., 2023).

Lineage specific gene analysis

Phylogenomic analysis was performed using Notung (Version 2.9) Command Line Interface (CLI). This software was used to

calculate the lineage-specific genes (LSGs) in the genus *Glycine*. The gene tree and species tree were reconciled under postfix species labels. Prior to reconciliation, the species tree was converted into a binary tree. Following the phylogenomic analysis, Notung saved the output in a Homolog table. This file contained descriptions of Paralogs, Orthologs, and Xenologs. The data provided by these results were then used to generate an UpSet plot using R. This plot indicated the conserved lineages across different species within the genus *Glycine*.

Comparative transcriptomics

To perform the expression analysis of identified NLR genes, we utilized available datasets from PRJNA628842 and PRJNA393826 bioprojects using the NCBI database (Supplementary Table S1). We aligned the raw read sequences using the reference genome of *G. max* and *G. soja* with HISAT (Pertea et al., 2015, 2016). Alignments were passed to StringTie for transcript assembly. Lastly, Ballgown was used to process the assembled transcripts and abundance to group the experimental conditions and identify the genes that were differently expressed between the conditions (Pertea et al., 2015, 2016).

Results

Expanded *NLR*ome in annuals in contrast to perennials species of *Glycine*

Significant differences of *NLR* genes were observed among annual and perennial species. Identified *NLR* genes can be classified into four sub-classes CC-*NLR*, CC_{G10}-*NLR*, TIR-*NLR* and CC_R-*NLR* (Figure 1A). *NLR* genes are often incomplete and have undergone pseudogenization due to gene duplication, retro-transposition, non-processed inactivation, and nonsense mutations. We further performed manual filtering for intact *NLR* genes from each class and a similar trend was identified among *Glycine* species (Supplementary Table S2) and 318 and 348 NLR genes were found in *G. soja and G. max*, respectively. Relatively contracted *NLR* gene distribution was found in perennials, including *G. falcata* (77), *G. dolichocarpa* (184), *G. syndetika* (111), *G. tomentella*, (99), *G. cyrtoloba* (93), *G. latifolia* (140), and *G. stenophita* (78) (Figure 1B). *NLR* genes belonging to sub-class CC-*NLR* and TIR-*NLR* have gained substantial expansions as compared to their distribution among seven perennials' species. Another subclass, CC_R-*NLR*, which is a key component in network of *NLR*ome. They remained highly conserved in all members of the family *Fabaceae* and their extent remained conserved, yet their expansion was also observed in annuals. Studying the factors contributing to the increased prevalence of *NLR* genes in annuals compared to perennial species is of immense importance. Recent phylogenomics suggest an increased number of non-redundant genes in annuals (129,006) as compared to perennials (109,827) (Zhuang et al., 2022). Expansion of *NLR* genes in annuals is also consistent with the super-pangenome of *Glycine* that provides evidence of a higher rate of non-core gene formation in annuals as compared to perennials. *G. dolichocarpa* subgenomes (A^t and D^t) have shown contraction of *NLR* genes upon comparison with their progenitor of *G. syndetika* (A) and *G. tomentella* (D). Significant reductions in TIR-*NLR* genes were observed in both A^t and D^t subgenomes. Interestingly *G. dolichocarpa-D*^{t} has shown increased ratio intact *NLR* genes as compared to its progenitor *G. tomentella* (D) in contrast to A^t which have shown considerable contraction as compared to *G. syndetika* (A) especially 66% TIR-*NLR* were lost after allopolyploidization. Overall, *NLR*ome in both A^t and D^t have shown biased subgenomic fractionation as D^t have shown higher ratio of intact *NLR* genes as compared to A^t. Comparing the A, D, A^t, and D^t genomes revealed that the A^t and D^t subgenomes of the allopolyploid lost 7,351 genes, with a higher number of losses from D (4,109) than from A (3,242) (Zhuang et al., 2022). This suggests that this biased fractionation of *NLR* genes is asymmetrical to the total number of genes lost in each subgenome.

NLR gene density is not a function of genome size

We further correlated the *NLR*ome of each species with respect to their genome size and examined the structural

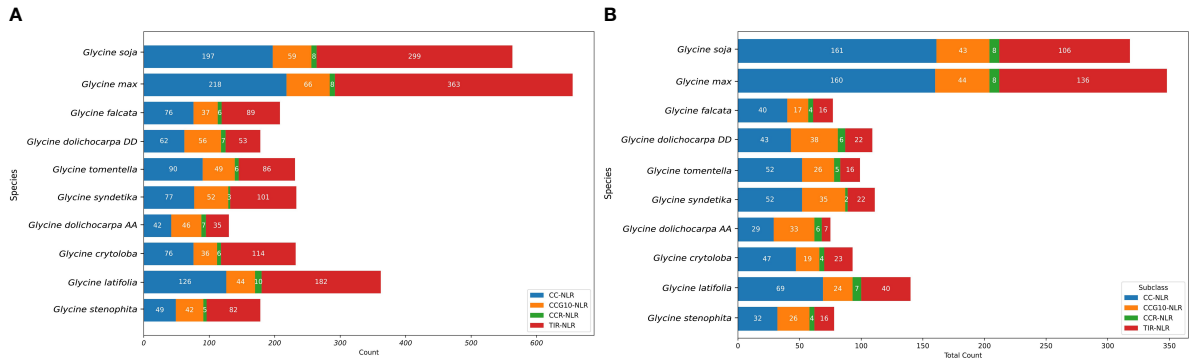


FIGURE 1
Comparative analysis of *NLR* genes and its subclasses in genus *Glycine:* A nucleotides domain-based perspective **(A)** Distribution of *NLR* genes in major classes across genus *Glycine* including partial and full length *NLR* genes **(B)** Distribution of Intact *NLR* Genes in *Glycine* Species.

organization of *NLR* genes. All genome assemblies anchored at the chromosomal level were analyzed to construct *NLR* gene density maps. A significant diversity of resistant genes was established across the different members of genus *Glycine*. The gene density map (Figure 2; Supplementary Table S3) demonstrates that the density of *NLR* genes does not correlate with genome size within the genus *Glycine*. For instance, *G. latifolia*, despite having a relatively smaller genome size of 995 MB, exhibits an expanded *NLR*ome. This contrasts with *G. falcata*, which has a larger genome size of 1.39GB and displays a reduced *NLR*ome density. Similarly, *G. cyrtoloba*, with a genome size of 1.3GB, has a constrained *NLR*ome. This is different from *G. syndetika* and *G. stenophita*, which, despite their smaller genome sizes of 948 and 940 MB, respectively, possess a higher *NLR*ome density than *G. cyrtoloba*. Interestingly, there is a contraction of *NLR* gene families in both subgenome A^t and D^t of the allotetraploid species *G. dolichocarpa* when compared to its ancestors, *G. tomentella* and *G. syndetika*. The contraction of *NLR* gene families in *G. dolichocarpa* suggests that polyploidization events do not necessarily lead to an expansion of gene families but can also result in their contraction. This observation aligns with previous findings that polyploidization can either increase or decrease certain gene families, contributing to the complexity of our understanding of genome evolution (Yin et al., 2019). These findings emphasize the complex dynamics of *NLR* gene density and its independence from genome size, highlighting the complex interaction of genetic factors in shaping the *NLR*ome across different species within the genus *Glycine*. The expanded *NLR*ome in *G. latifolia* and the reduced *NLR*ome in *G. falcata*, despite their contrasting genome sizes, indicate that *NLR* gene density is not merely a function of genome size. Instead, it may be influenced by a multitude of factors, including the species' evolutionary history, environmental pressures, and genetic mechanisms. We further illustrated that individual gene density maps on each chromosome revealed significant variation in the distribution of *NLR* genes. Annual species like *G. max* and *G. soja* have more pronounced high-density regions (Supplementary Figure S1), suggesting a higher concentration of NLR genes, which could be crucial for their immune response. In contrast, perennial species show a more uniform and lower density distribution, indicating fewer NLR genes or a more even spread across their genomes.

Classification of identified *NLR* genes

Furthermore, a comprehensive classification of *NLR* genes was performed across the genus *Glycine*. The evolutionary history of *NLR* genes in the genus *Glycine* can be traced by constructing their phylogenetic relationships. This is achieved by analyzing the amino acid sequences of the conserved NB-ARC domain. The TNL clade branched out as expected. The CNL class split into three major subclades: CCR-*NLR*, CC-*NLR*, and CCG10-*NLR*. Within the CC-*NLR* sub-clade, we identified four significant sub-groups: CNL-Un, CNL-G11, CNL-G7, and G4. The TNL clade exhibited polyphyletic traits, suggesting multiple radiations and significant diversification (Figure 3). In comparison, CC_{G10}-CNL exhibited considerable diversity and expansion with seven radiations. CC_R-*NLR* doesn't diversify and remains highly conserved across all species. The complete absence of groups from G1-G8 across the genus *Glycine* aligns with previous findings (Asif et al., 2023, Rani et al., 2023, Areej et al., 2023). The multiple radiations of G7 and G4 also suggest their significant diversification throughout the genus *Glycine*. The expansion and

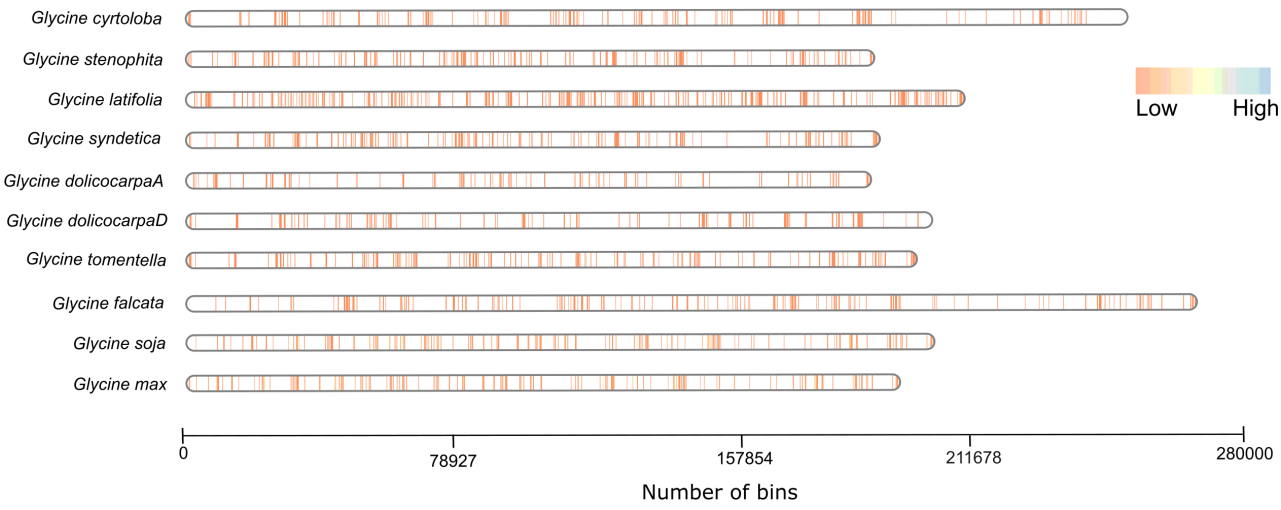


FIGURE 2
Gene density map: This graph displays bins on the x-axis, each representing a 5kb segment of the genome. The color gradient from brown to blue signifies the varying density of *NLR* genes. This graphical representation provides a quantitative correlation between *NLR* gene density and genomic size of different species of *Glycine*.

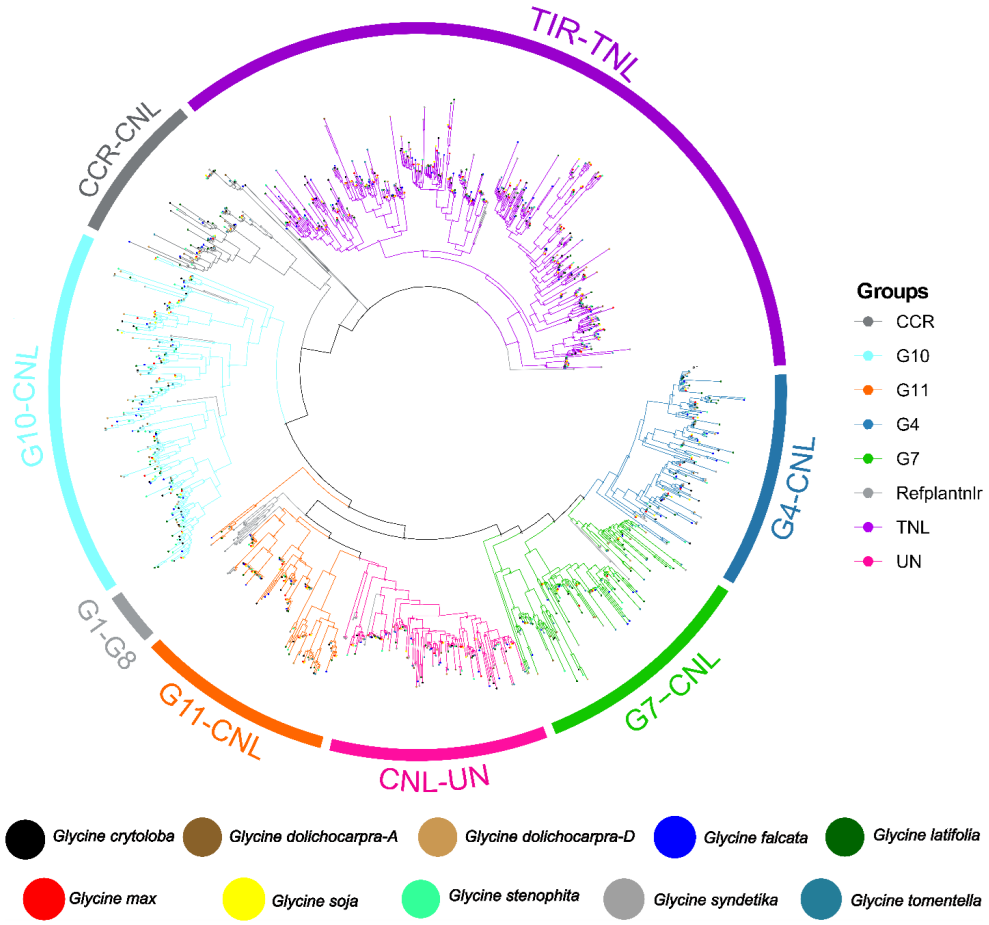


FIGURE 3
Phylogenetic distribution of 8 genomes and two subgenomes of genus *Glycine* among four major subclasses of *NLR* genes 1) CC-*NLR* 2) CC_{G10}-*NLR* 3) CC_R-*NLR* 4) TIR-*NLR*.

absence of certain groups in genus *Glycine* directly related to the specific pathogen they detect. The encounters of *Glycine* species with different pathogens lead to the diversification of certain groups like we observe in G10, G4 and G7, and the absence of certain pathogens also leads to the lack of associated sub-groups, as in this case G1-G8.

Contrast in gene birth and lost ratio in annuals and perennials

Since we have observed that the annual lineage (*soja*) of *Glycine* have an expanded *NLR*ome as compared to the perennial lineage (*wild*). These expansions in the *NLR*ome are either due to the birth of new genes or duplication of existing genes. The study of gene gain and loss of *NLR* genes in the genus *Glycine* was conducted using Orthofinder and CAFÉ analysis. Family trees for each *Glycine* species and its subgenome were constructed by comparing them with *Phaseolus vulgaris*, its closest allies in the legume tribe. Despite the unique whole genome duplication (WGD) event that occurred in the ancestor of annuals (*soja*) and perennials (wild), a significant loss of *NLR* genes was observed in the ancestral lineage (Figure 4). The contraction of *NLR* genes after WGD in *Glycine* is consistent with previous observation that the WGD event is followed by a trend of diploidization leading to contraction. Furthermore, following the divergence from perennials, the annual lineage (*Soja*), exhibited an expansion of *NLR* gene families (Figure 4). The most significant gene gain of 42 gene families were observed in the ancestral lineage of the subgenus *Soja*. The highest rate of terminal duplication is found in both annual species of the subgenus *Soja*. In contrast, the perennial subgenus *G. wild* demonstrates an overall decrease in the number of *NLR* gene families across the subgenus. The greatest gene loss was observed in G. *falcata*, which lost 39 gene families and gained only 6 of them. Terminal duplication also shows a downward trend across the perennials,except for *G. latifolia*, which gained twenty-nine gene families and lost only seven of them. Considering the evidence from gene birth and death analysis, we can hypothesize that the birth of new sub-gene families, lineage specific and terminal duplication could be the major reason for expanded *NLR*ome in annuals species. We also compared the shared orthologs between perennials and *G. max* (Figures 4B, C). In total 49 shared orthologs were present between perennials diploid

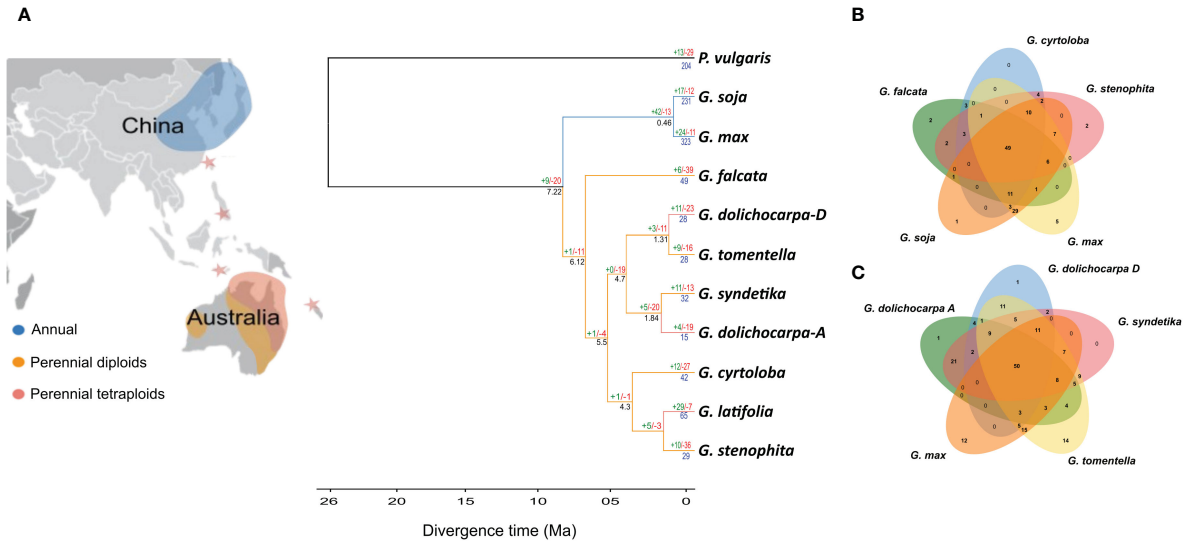


FIGURE 4
Unraveling gene gains and losses in *Glycine* species: **(A)** The phylogenetic tree visualizes the evolutionary trajectory of *NLR* genes, highlighting gene gain (green), loss (red), and duplication (blue) events across the chronological spectrum. Additionally, the branch color indicates the global distribution of each species, linking genetic evolution to geographical occurrence. **(B, C)** The Venn diagram illustrates the distribution of species-specific and orthologous *NLR* genes across the sub-genus, providing insights into the genetic diversity and evolutionary relationships among the species.

species (*G. falcata, G. cyrtoloba and G. stenophita*: 4B). Increased number of shared ortholog of up-to 59 were present between *G. max* and polyploid perennials (*G. dolichocarpa, G. tomentella* and *G. syndetika*:4C).

Evolutionary history of *NLR* gene evolution

Considering the immense importance of gene duplication, as illustrated by gene gain and loss analysis. We explored the duplication history of *Glycine* by comparing *Ks* values between paralogs of each subgroup. The closest estimates suggest that divergence between annuals and perennials occurred at ~7 Mya. Collective Ks values obtained from all members of the genus *Glycine* have shown a common duplication curve since 26 Mya after the split from the recent common ancestor *Phaseolus vulgaris*. Substitution analysis further suggests that the rate of duplication increased after the *Glycine*-specific whole-genome duplication (WGD) event, which occurred approximately ~10 Mya (Figure 5). Despite contraction after ongoing diploidization, the perennials lineage underwent a high ratio of gene duplication between an estimated time range of 3–8 Mya (*Ks*= 0.05–0.01) followed by gradual reduction at the end of this marked increases in the duplication of CC_{G10}-*NLR* and TIR-*NLR* were observed during this accelerated gene duplication cycle. However, annuals continued their gene duplications until recently, and the highest duplication rate was observed between 0.1–0.5 Mya. This suggests that annuals species of *Glycine* remained expanding their *NLR*ome and the highest duplication occurred after the divergence of *G. max* and *G. soja* (annuals speciation ~ 0.47 Mya). Interestingly, both annuals' species *G. max* and *G. soja* have showed pronounced duplications of G4-CNL subgroup that ultimately led to diversification and expansion of CC-*NLR* genes. In short, recent gene duplication is responsible for the expansion of the *NLR*ome in annuals. Secondly, the duplication assay does not provide evidence of recent and accelerated gene duplication rates in subgenomes *G. dolichocarpa*, which is consistent with lower terminal duplication rates provided by Orthofinder analysis. It further suggests that the D^t-subgenome of *G. dolichocarpa* could have expanded through processes other than gene duplication, such as recombination and transposition.

Conservation of lineage and species-specific gene

We further evaluated the conservation of lineage-specific gene across genus *Glycine* by comparing *NLR* gene tree and species tree using Notung tool. The highest conservation of lineage-specific genes was found among annuals (*G. max* and *G. soja*) where 64 both species shared 64 orthologs (Figure 6). Among perennials highest lineage conservation were found among *G. latifolia* and *G. falcata* which is contrast with their relatively earlier divergence (~6 Mya). Furthermore, *G. falcata* also showed a significant share of common orthologs with all members of perennials, suggesting the highest conservation of ancestral *NLR* genes. Comparison of perennials and annuals revealed ten or less than ten conserved lineages. This significant lack of conserved lineages can be explained by the geographical isolation of perennials species since most perennials are native to Australia and annuals are native to eastern China. In the case of tetraploid *G. dolichocarpa*, both subgenomes A^t and D^t have shared 24 and 26 *NLR* genes lineages with their progenitors *G. syndetika* (A) and *G. tomentella* (D). The highest number of species-

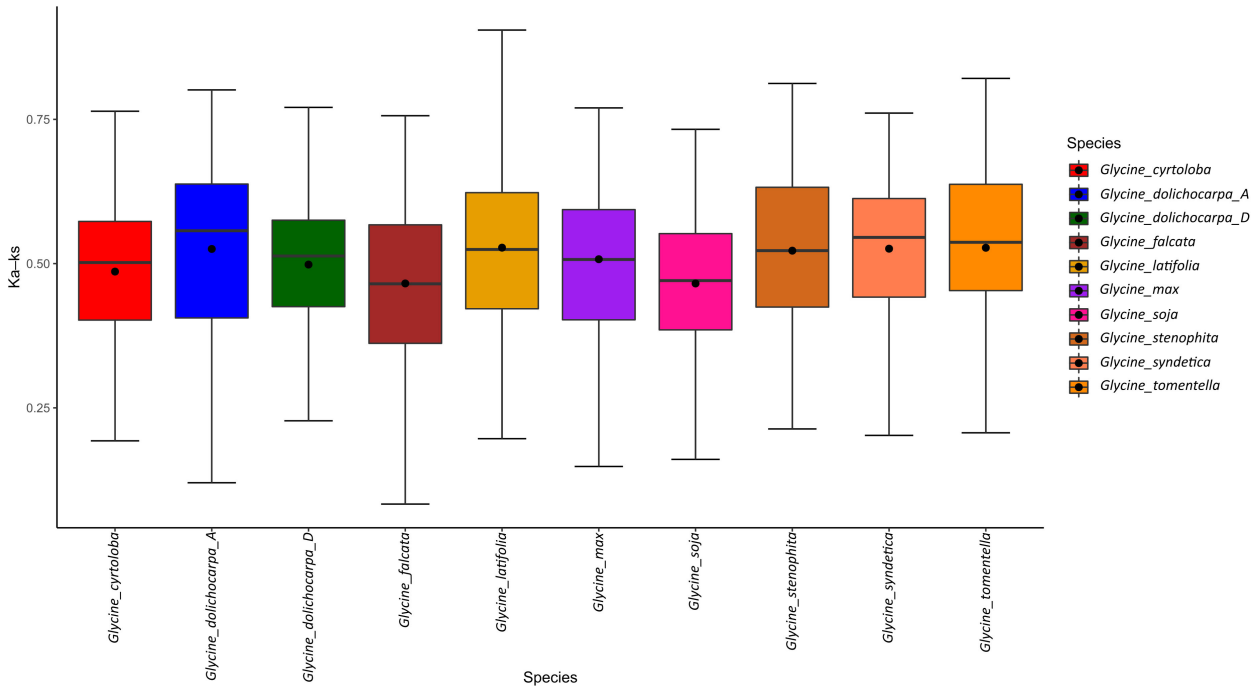


FIGURE 5
Evolutionary history of *NLR* gene duplication in *Glycine* species.

specific lineages was observed in *G. latifolia*, which is consistent with the rapid birth of genes and terminal duplication discussed earlier (Figure 4). Overall, a higher ratio of species-specific genes was found in perennials and annuals, demonstrating less complex specific *NLR* genes repertoire.

Synteny analysis

Considering the highest degree of lineage-specific genes among annuals and lower conservation among perennials, we further performed an in-depth comparison by using synteny analysis. It provides a deep understanding of the evolutionary connections, genomic modifications, and preserved functions across various organisms. Among annuals, a considerable amount of syntenic links were discovered among these species (Figure 3). This suggests a high level of genomic conservation between *G. max* and G. *soja*, which is expected given their close evolutionary relationship within the annuals. *G. soja* and *G. max* have shown the greatest number of ortholog clusters throughout the chromosomes (Figure 7A). We further compared annuals (*G. max* and *G. soja)* and perennial member *G. latifolia. G. max* and *G. soja* exhibit a significant degree of synteny between them, which contrasts with *G. latifolia* (Figure 7B). The latter demonstrates a declining trend in the syntenic relationship between orthologs within the subgenus *Soja*. Further, *G. soja* was substituted with the perennial *G. dolichocarpa*. Interestingly, the members of the subgenus G. *wild* did not exhibit as much synteny among themselves as we found within the subgenus *Soja* (Figure 7C). Major clusters were identified on chromosomes 1, 2, 3, 18 and 20. High syntenic relation between *G. max* and G. *soja* indicate that they both share a common ancestor, and a large part of the genome has been maintained over time, which shows high genomic stability.

Comparative transcriptomics of identified *NLR* genes

We further evaluated the expression identified *NLR* genes using available datasets. We performed a comparative transcriptomic analysis of both *G. soja* and *G. max* in response to infection by *Fusarium oxysporum* (Figure 8). Within the 45-day duration, a total of 45 genes were expressed in *G. soja*, whereas 53 genes were expressed in *G. max*. These genes were categorized into distinct groups: CCR, G10, G11, TNL, and CNL-UN. Notably, the expression of *NLR* genes was significantly higher in *G. soja* (wild soybean) compared to *G. max* (cultivated soybean). Specifically, under both infected and non-infected conditions in the wild, genes such as Glysoja.17G046920.1, Glysoja.14G038028.1, Glysoja.17G046921.1, Glysoja.05G011704.3, and Glysoja.14G038027.1 exhibited elevated expression levels in infected conditions, whereas their expression was comparatively lower in non-infected conditions. These highly expressed genes primarily belonged to the CC_R-*NLR* and TIR-*NLR* groups. Contrastingly, *G. max* has shown opposite expression pattern. *NLR* gene expression was higher under non-infected conditions compared to infected conditions. Genes such as Glyma.16G215100.1, Glyma.14G079600.1, Glyma.17G245500.1, Glyma.17G245600.1, Glyma.17G179200.1, Glyma.05G082200.1, and Glyma.16G137200.1, categorized under CC_R-*NLR* and TIR-*NLR* groups, demonstrated this differential expression pattern.

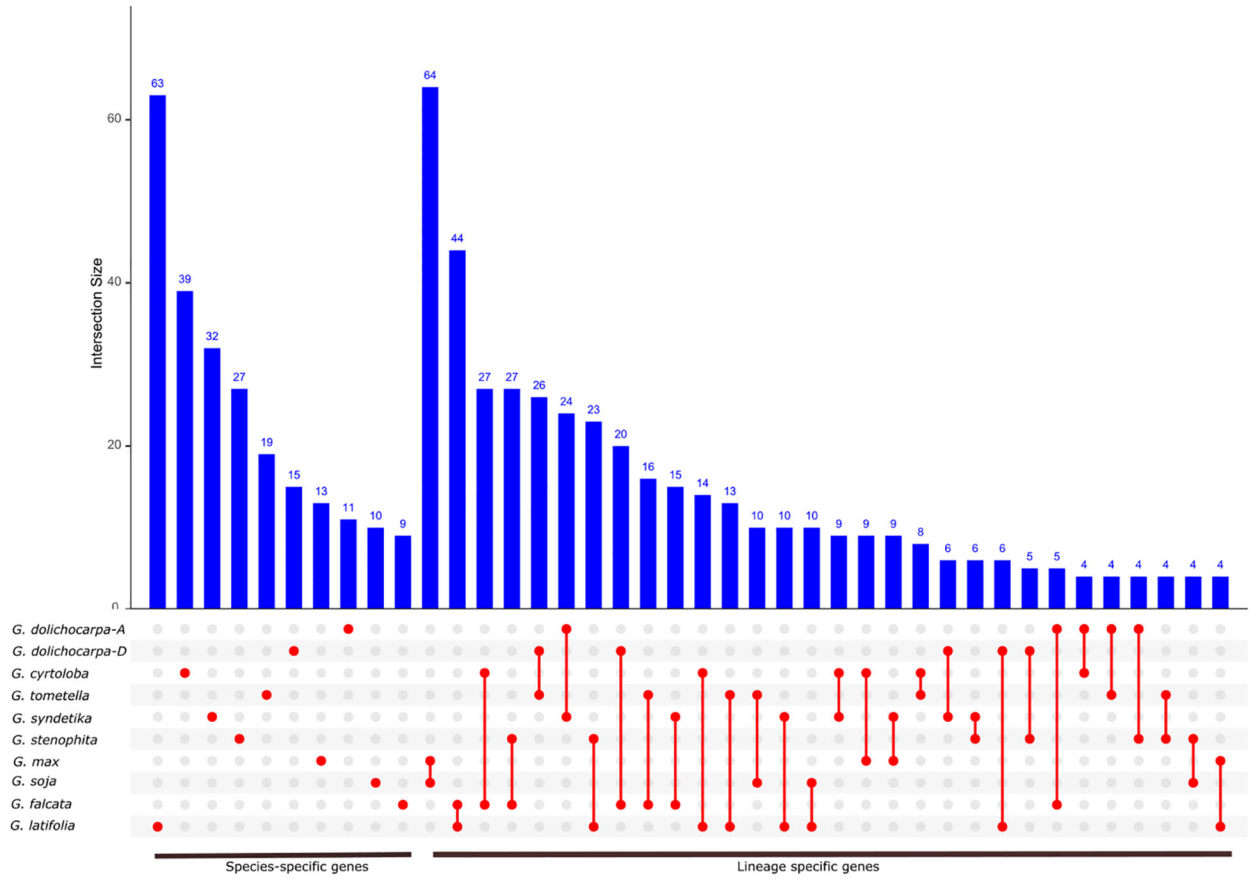


FIGURE 6
Identification of species and lineage specific *NLR* genes in genus *Glycine*: The blue bars denote the count of lineage-specific genes within each species. The red dots and lines below the bars highlight the lineages that are conserved across multiple species.

Additionally, Glyma.14G079500.1 and Glyma.16G210600.1 exhibited notably high expression levels, specifically in non-infected conditions. Overall, the analysis indicates that the expression of *NLR* genes is substantially higher in wild soybean species compared to cultivated soybeans in response to *Fusarium oxysporum*. Furthermore, distinct expression patterns were observed between the two species under both infected and non-infected conditions, with different sets of genes displaying varied expression levels in response to the pathogen.

In the PRJNA628842 dataset, the seeds of two lines of *G. max* (LHY and WT) were grown in the Hogland Solution for two weeks and these samples were collected according to the ZT (zeitgeber time) (Figure 9). LHY is a transgenic line, that overexpresses Late Elongated Hypocotyl gene (LHY) which is responsible for the circadian movement of leaf in plants (Wang et al., 2021). We performed comparative transcriptomics between two lines of *G. max* LHY and WT and their response to drought stress (LHY-D and WT-D). In normal conditions relatively lower expression of 25 *NLR* genes were observed that can be classified into 5 groups (CCR, G10, G11, TNL and UN), highest number of genes that were expressed belonged to helper *NLR* (CC_R-*NLR*). However, under drought conditions in LHY and WT, the expression rate of *NLR* genes were higher as compared to well-watered conditions. In total 7 genes have higher rate of expression in both LHY-D and WT-D condition belonging to CC_R-*NLR*, TIR, and G11-CNL subgroups.

Discussion

Soybean, originally from China, is now the world's most widely grown oil and protein seed crop. Despite its agricultural importance, the primary gene pool of soybean, which includes *G. max* (cultivated soybean) and *G. soja* (wild soybean), exhibits low genetic diversity. This limited diversity is a significant constraint on the crop's environmental resilience and yield potential. *G. max* is particularly susceptible to major pathogens, including viruses, bacteria, and fungi (Araújo et al., 2019; Bandara et al., 2020). Nucleotide-binding site leucine-rich repeat (NLR) genes play a crucial role in plant immunity by recognizing pathogen effectors and triggering defense responses.

Expansion of NLRome in annuals

Previous study for the characterization of NLR genes was limited to annual species (*G. max* and *G. soja*) and due to

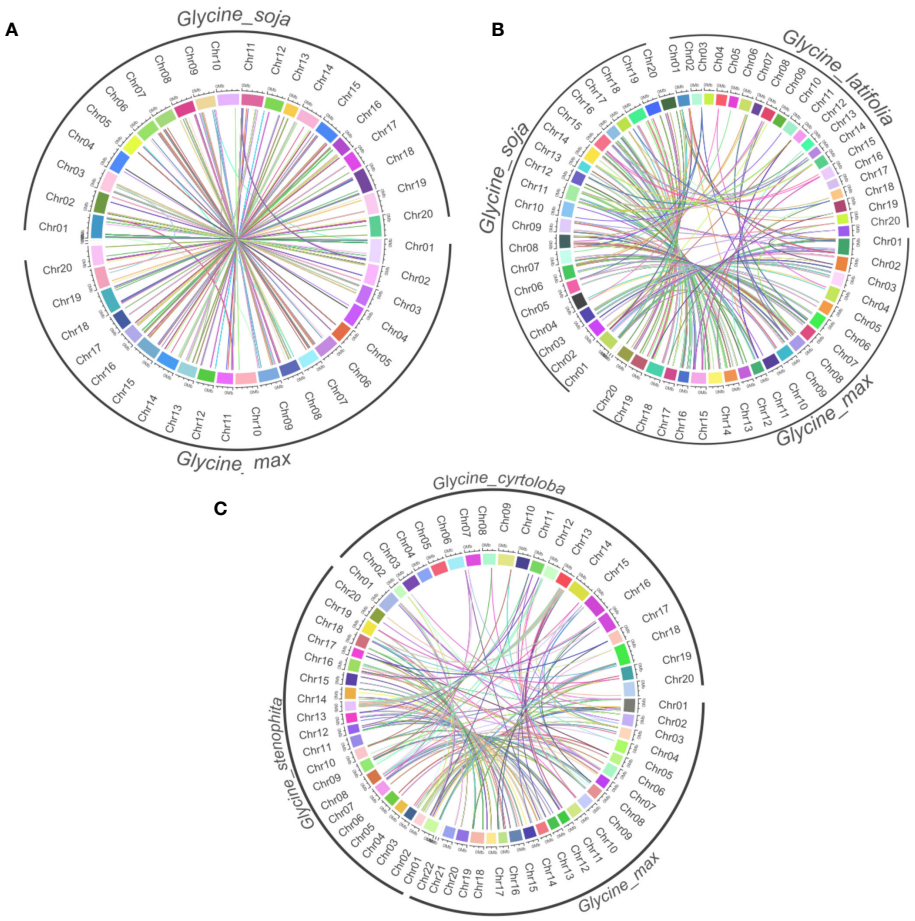


FIGURE 7
Synteny analysis of *NLR* genes among annuals and perennials species. Each panel represents syntenic relations among **(A)** annuals (*G. max* and *G. soja*) **(B)** Annuals species and *G. latifolia* **(C)** *G. dolichocarpa*, *G. max* and *G. latifolia*..

fragmented genome assemblies limited number of NLR gene were identified. For example, reduced NLR were identified from *G. max* (Zheng et al., 2016; Afzal et al., 2022). Utilization of updated reference genome and mining approaches has allowed detailed understanding of NLR gene evolution. Our study reveals that both *G. max and G. soja* have an expanded repertoire of NLR genes compared to their perennial relatives. This expansion is attributed to recent gene duplications occurring between 0.1 to 0.5 million years ago in their common ancestor and after their speciation. This is evidenced by a higher ratio of duplicate NLR

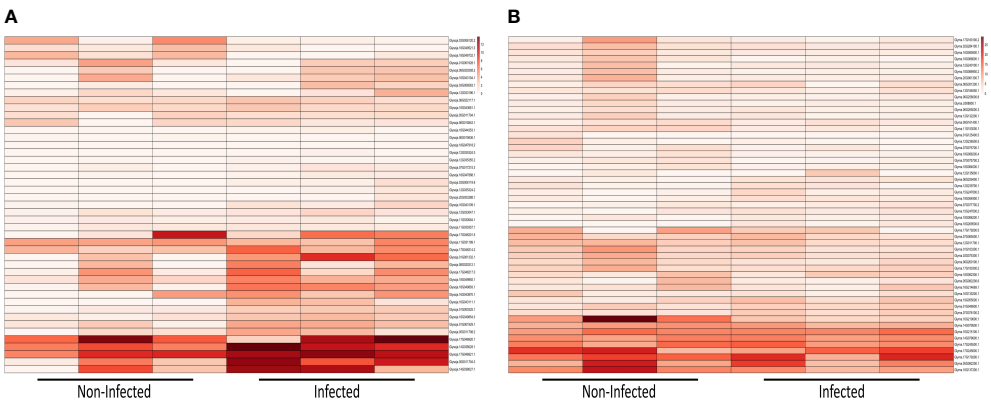


FIGURE 8
Expression analysis upon *Fusarisum oxysporum* infection **(A)** heatmap of *NLR* genes expressions in infected and non-infected *G. soja* **(B)** heatmap of *NLR* genes expressions in infected and non-infected *G. max*.

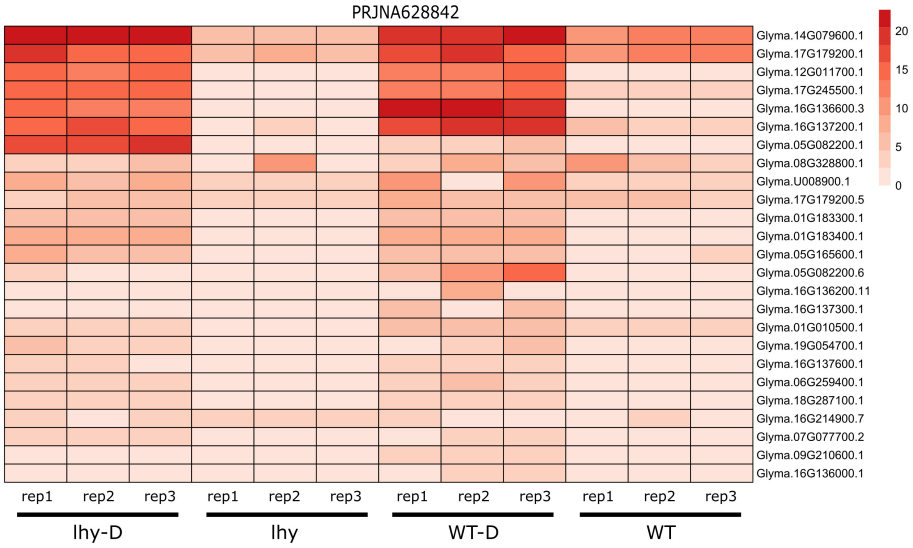


FIGURE 9
Expression analysis of *NLR* genes upon drought treatment in wild and transgenic *G. max*. LHY lines overexpression LHY genes illustrates higher expression of *NLR* genes drought conditions.

genes to singletons and the rapid emergence of non-core genes in annuals in annual species compared to perennials (Liu et al., 2018; Zhuang et al., 2022).

Contraction in Perennials

Perennial species have experienced significant contraction of NLR genes following a whole-genome duplication event approximately 10 million years ago. This contraction has led to a more diversified but smaller set of NLR genes, likely due to gene duplications that occurred between 4 to 7 million years ago after diverging from the annual lineage. It is consistent with previous studies that identified multiple soybean cyst nematode (SCN) population in *G. tomentella* and other species *G. argyria* and *G. pescadrensis* also showed resistance to all tested SCN populations (Wen et al., 2017). Similarly perennial *Glycine* species (*G. argyrea, G. clandestina, G. dolichocarpa, G. tomentella and G. canescens)* have demonstrated resistance to soybean rust (*Phakopsora pachyrhizi*) (Herman et al., 2020).

Genomic stability in annuals

The high syntenic relationship between *G. max and G. soja* indicates that a large part of their genome has been maintained over time, demonstrating high genomic stability in the annual lineage. Synteny analysis revealed a high degree of genomic conservation between the annual species *G. max and G. soja*, indicating their close evolutionary relationship and shared ancestry. This is consistent with recently constructed *De-novo* assembled pan genome of soybean wild relatives that provide evidence of greater genomic stability of *G. max* and *G. soja* as compared to perennial species and confirmed that large part of genome has been maintained over time (Joshi et al., 2013). In contrast, the perennial species *G. latifolia* exhibited a declining trend in syntenic relationship with the annuals, suggesting divergence within the subgenus Soja. Previous study has also decreased trend of syntenic relation as only 12 out of 12 G. latifolia linkage groups were identified that were colinear with *G. max* chromosomes (Chang et al., 2014). Within the perennial subgenus *Glycine* wild, the species did not exhibit as much synteny among themselves as observed within the annual subgenus Soja. Major syntenic clusters were identified on chromosomes 1, 2, 3, 18, and 20.

Unique NLR Repertoire in *G latifolia*

G. latifolia, known for its high levels of resistance to multiple soybean pathogens and pests, encodes a unique repertoire of NLR genes that are highly species-specific. This diversification might have occurred after the common duplication curve between 4–7 Mya. Despite the ongoing diploidization trend in perennial species, *G. latifolia* has gained 29 gene families with an accelerated terminal duplication rate relative to the closest species (*G. cyrtoloba* and *G. stenophita*). Gene gain and loss analysis provide strong evidence that the birth of new gene families and terminal duplication are significant reasons for the highly divergent evolution of NLR genes in *G. latifolia*. Previous studies highlighted the presence of 3,148 unique sets of genes and noted the overrepresentation of NB-ARC encoding genes (Zhuang et al., 2022). Similarly, comparative analysis with five legume species showed that genes related to defense responses were significantly overrepresented in *Glycine*-specific orthologous gene families (Liu et al., 2018).

Gene birth and losses is a significant driver of divergent evolution

The evolutionary history of NLR genes in soybean annuals and perennials plants indicates that gene duplications and losses have played a significant role in shaping their current NLR gene profiles. The study of gene gain and loss of NLR genes in the genus *Glycine* revealed that annual species have a higher rate of gene birth and terminal duplication compared to perennials. The most significant gene gain was observed in the ancestral lineage of the subgenus *Soja*, while the greatest gene loss was observed in *G. falcata*. *G. latifolia*, however, showed a significant gain of gene families, indicating a dynamic evolutionary process. Recently constructed high density linkage maps for *G. latifolia* and their comparison with G. *max* has found significant chromosomal rearrangements in perennial species and annual species *G. max* has undergone more frequent gene duplication contributing to their genetic diversity adaptability (Chang et al., 2014). These dynamics are crucial for understanding how plants adapt to pathogen pressures and environmental changes.

Conservation of lineage and species-specific genes

The study found the highest conservation of lineage-specific genes among annuals (*G. max and G. soja*), while perennials showed a higher ratio of species-specific genes. This significant lack of conserved lineages between annuals and perennials can be explained by their geographical isolation and different evolutionary pressures.

Effect of allopolyploidy on NLR evolution

This study provides new insights into the effect of allopolyploidy on the evolution of NLR genes. The availability of the complete genome of A and D diploid progenitors allows a precise definition of its sub-genome and its genome-wide distribution of NLR genes. Comparing the A, D, A^t, and D^t genomes revealed that the A^t and D^t sub-genomes of the allopolyploid lost 7,351 genes, with a higher number of losses from D (4,109) than from A (3,242) (Zhuang et al., 2022). Conversely, NLR distribution was asymmetrical to the already described complete gene fractionation pattern. Marked expansion of NLR gene distribution was observed for D^t as compared to At. This asymmetric expansion of the NLRome in the subgenomes of *G. dolichocarpa* could possibly be due to homologous sequence exchanges (HSEs) and illegitimate recombination. The availability of chromosomal-anchored genome sequences of additional polyploids and their progenitors will further provide a better understanding of the role of polyploids on the evolution of NLR genes.

Implications for soybean breeding

The findings highlight the importance of leveraging the genetic diversity within the *Glycine* genus to improve soybean's disease resistance. By identifying and introgressing beneficial NLR genes from wild relatives and other sources, breeders can develop soybean varieties with enhanced resilience and yield potential. This study has significant implications for crop breeding and development of disease-resistance soybean cultivars. The phylogenetic distribution of NLR genes provides a catalog of genetic resources that can be used for crop improvement. By utilizing diverse *Glycine* germplasm, breeders can introduce new resistance genes into elite cultivars, enhancing genetic diversity and crop resilience. The conserved NLR genes are valuable resource for developing broad-specturm disease resistance. G4, G7, G11, CC_R-CNL and CCG10-CNL perennial lineage-specific NLR genes of *Glycine* offer an opportunity to develop unique disease resistance traits tailored to regional challenges. These genes can be incorporated into breeding programs to create cultivars with enhanced or novel resistance traits. The use of genomic and transcriptomic resources has enabled researchers to uncover the mechanisms underlying the diversification and maintenance of *NLR* gene repertoires, providing valuable insights for future studies on plant immune system evolution and disease resistance. The provided results offer valuable insights into the genetic diversity, evolution, and regulation of *NLR* genes in related plant species, supporting the findings of the study on the genus *Glycine*. These results underscore the broader significance of *NLR* gene research in the context of plant immunity and evolutionary biology.

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

AS: Data curation, Formal Analysis, Methodology, Software, Writing – original draft, Writing – review & editing. HN: Formal Analysis, Funding acquisition, Project administration, Resources, Visualization, Writing – original draft, Writing – review & editing. FS: Project administration, Resources, Software, Writing – original draft, Writing – review & editing. SN: Data curation, Formal Analysis, Methodology, Writing – original draft, Writing – review & editing. MD: Data curation, Funding acquisition, Methodology, Resources, Software, Writing – original draft, Writing – review & editing. RI: Data curation, Formal Analysis, Methodology, Validation, Writing – original draft, Writing – review & editing. UM: Data curation, Methodology, Software,

Visualization, Writing – original draft, Writing – review & editing. AA: Data curation, Investigation, Methodology, Resources, Writing – original draft, Writing – review & editing. NA: Data curation, Funding acquisition, Methodology, Resources, Validation, Visualization, Writing – original draft, Writing – review & editing. RM: Data curation, Methodology, Software, Writing – original draft, Writing – review & editing. SSh: Data curation, Methodology, Writing – original draft. MR: Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing. SSe: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was funded by Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2024R62), Princess Nourah bint Abdulrahman University, Riyadh, 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.

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2024.1383135/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Gene density map with respect to each chromosome of *Glycine* species (ranged 1–20). The gradient color is provided to represent the gene density of each chromosome with respect to number bins present in the chromosome.

References

Afzal, M., Alghamdi, S. S., Nawaz, H., Migdadi, H. H., Altaf, M., El-Harty, E., et al. (2022). Genome- wide identification and expression analysis of CC-NB-ARC-LRR (NB-ARC) disease-resistant family members from soybean (Glycine max L.) reveal their response to biotic stress. *J. King Saud Univ. - Sci.* 34. doi: 10.1016/j.jksus.2021.101758

Andersen, E. J., Ali, S., Byamukama, E., Yen, Y., and Nepal, M. P. (2018). Disease resistance mechanisms in plants. *Genes* 9, 339. doi: 10.3390/genes9070339

Araújo, A. C., Fonseca, F. C. D. A., Cotta, M. G., Alves, G. S. C., and Miller, R. N. G. (2019). Plant NLR receptor proteins and their potential in the development of durable genetic resistance to biotic stresses. *Biotechnol. Res. Innov.* 3, 80–94. doi: 10.1016/j.biori.2020.01.002

Areej, A., Nawaz, H., Aslam, I., Danial, M., Qayyum, Z., Rasool, U. A., et al. (2023). Investigation of *NLR* genes reveals divergent evolution on *NLR*ome in diploid and polyploid species in genus trifolium. *Genes (Basel)* 14, 867. doi: 10.3390/genes14040867

Asif, J., Qureshi, F., Zain, M., Nawaz, H., Naz, E., Fareed, S., et al. (2023). Investigation of resistance genes in genus vigna reveals highly variable *NLR*ome in parallel domesticated member species. *Genes (Basel)* 14, 1129. doi: 10.3390/genes14061129

Bandara, A. Y., Weerasooriya, D. K., Bradley, C. A., Allen, T. W., and Esker, P. D. (2020). Dissecting the economic impact of soybean diseases in the United States over two decades. *PloS One* 15. doi: 10.1371/journal.pone.0231141

Calle García, J., Guadagno, A., Paytuvi-Gallart, A., Saera-Vila, A., Amoroso, C. G., D'esposito, D., et al. (2022). PRGdb 4.0: an updated database dedicated to genes involved in plant disease resistance process. *Nucleic Acids Res.* 50, D1483–D1490. doi: 10.1093/NAR/GKAB1087

Chang, S., Thurber, C. S., Brown, P. J., Hartman, G. L., Lambert, K. N., and Domier, L. L. (2014). Comparative mapping of the wild perennial Glycine latifolia and soybean (G. max) reveals extensive chromosome rearrangements in the genus Glycine. *PloS One* 9. doi: 10.1371/journal.pone.0099427

Cui, H., Tsuda, K., and Parker, J. E. (2015). Effector-triggered immunity: from pathogen perception to robust defense. *Annu. Rev. Plant Biol.* 66, 487–511. doi: 10.1146/annurev-arplant-050213-040012

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26, 2460–2461. doi: 10.1093/bioinformatics/btq461

Egan, A. N., and Doyle, J. (2010). A comparison of global, gene-specific, and relaxed clock methods in a comparative genomics framework: dating the polyploid history of soybean (*Glycine* max). *Syst. Biol.* 59, 534–547. doi: 10.1093/sysbio/syq041

Emms, D. M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 20. doi: 10.1186/s13059-019-1832-y

Forrester, N. J., and Ashman, T. L. (2018). The direct effects of plant polyploidy on the legume-rhizobia mutualism. *Ann. Bot.* 121, 209–220. doi: 10.1093/aob/mcx121

Hao, Z., Lv, D., Ge, Y., Shi, J., Weijers, D., Yu, G., et al. (2020). RIdeogram: drawing SVG graphics to visualize and map genome-wide data on the idiograms. *PeerJ Comput. Sci.* 6, 1–11. doi: 10.7717/peerj-cs.251

Herman, T. K., Han, J., Singh, R. J., Domier, L. L., and Hartman, G. L. (2020). Evaluation of wild perennial Glycine species for resistance to soybean cyst nematode and soybean rust. *Plant Breed.* 139, 923–931. doi: 10.1111/pbr.12834

Innes, R. W., Ameline-Torregrosa, C., Ashfield, T., Cannon, E., Cannon, S. B., Chacko, B., et al. (2008). Genome analysis differential accumulation of retroelements and diversification of NB-LRR disease resistance genes in duplicated regions following polyploidy in the ancestor of soybean 1[W][OA]. *Plant Physiol.* 148, 1740–1759. doi: 10.1104/pp.108.127902

Jones, J. D. G., Vance, R. E., and Dangl, J. L. (2016). Intracellular innate immune surveillance devices in plants and animals. *Sci. (1979)* 354. doi: 10.1126/science.aaf6395

Joshi, T., Valliyodan, B., Wu, J. H., Lee, S. H., Xu, D., and Nguyen, H. T. (2013). Genomic differences between cultivated soybean, G. max and its wild relative G. soja. *BMC Genomics*, 14. doi: 10.1186/1471-2164-14-S1-S5

Kim, K., El Baidouri, M., Abernathy, B., Iwata-Otsubo, A., Chavarro, C., Gonzales, M., et al. (2015). A comparative epigenomic analysis of polyploidy-derived genes in soybean and common bean. *Plant Physiol.* 168, 1433–1447. doi: 10.1104/pp.15.00408

Kourelis, J., Sakai, T., Adachi, H., and Kamoun, S. (2021). RefPlant*NLR* is a comprehensive collection of experimentally validated plant disease resistance proteins from the *NLR* family. *PloS Biol.* 19. doi: 10.1371/journal.pbio.3001124

Li, K.B. (2003). ClustalW-MPI: ClustalW analysis using distributed and parallel computing. *Bioinformatics* 19, 1585–1586. doi: 10.1093/bioinformatics/btg192

Li, Y. H., Zhou, G., Ma, J., Jiang, W., Jin, L. G., Zhang, Z., et al. (2014). *De novo* assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits. *Nat. Biotechnol.* 32, 1045–1052. doi: 10.1038/nbt.2979

Liu, Q., Chang, S., Hartman, G. L., and Domier, L. L. (2018). Assembly and annotation of a draft genome sequence for *Glycine* latifolia, a perennial wild relative of soybean. *Plant J.* 95, 71–85. doi: 10.1111/tpj.13931

Manzoor, A., Ahmad, T., Bashir, M. A., Hafiz, I. A., and Silvestri, C. (2019). Studies on colchicine induced chromosome doubling for enhancement of quality traits in ornamental plants. *Plants (Basel)* 8. doi: 10.3390/plants8070194

Nguyen, L. T., Schmidt, H. A., Von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300

Pan, Q., Wendel, J., and Fluhr, R. (2000). Divergent evolution of plant NBS-LRR resistance gene homologues in dicot and cereal genomes. *J. Mol. Evol.* 50, 203–213. doi: 10.1007/s002399910023

Pertea, M., Kim, D., Pertea, G. M., Leek, J. T., and Salzberg, S. L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat. Protoc.* 11, 1650–1667. doi: 10.1038/nprot.2016.095

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T. C., Mendell, J. T., and Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* 33, 290–295. doi: 10.1038/nbt.3122

Quinlan, A. R., and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. doi: 10.1093/bioinformatics/btq033

Qureshi, F., Mehmood, A., Khan, S. A., Bilal, M., Urooj, F., Alyas, M., et al. (2023). Macroevolution of *NLR* genes in family Fabaceae provides evidence of clade specific expansion and contraction of *NLR*ome in Vicioid clade. *Plant Stress* 10, 100254. doi: 10.1016/j.stress.2023.100254

Rani, S., Zahra, R., Bakar, A., Rizwan, M., Sultan, A.-B., Zain, M., et al. (2023). Dynamic Evolution of NLR Genes in Dalbergioids. *Genes* 14, 377. doi: 10.3390/GENES14020377

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

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

Seo, E., Kim, S., Yeom, S. I., and Choi, D. (2016). Genome-wide comparative analyses reveal the dynamic evolution of nucleotide-binding leucine-rich repeat gene family among solanaceae plants. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.01205

Sherman-Broyles, S., Bombarely, A., Powell, A. F., Doyle, J. L., Egan, A. N., Coate, J. E., et al. (2014). The wild side of a major crop: Soybean's perennial cousins from Down Under. *Am. J. Bot.* 101, 1651–1665. doi: 10.3732/AJB.1400121

Suyama, M., Torrents, D., and Bork, P. (2006). PAL2NAL: Robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* 34. doi: 10.1093/nar/gkl315

Wang, J., and Chai, J. (2020). Molecular actions of *NLR* immune receptors in plants and animals. *Sci. China Life Sci.* 63, 1303–1316. doi: 10.1007/s11427-019-1687-6

Wang, W., Chen, L., Fengler, K., Bolar, J., Llaca, V., Wang, X., et al. (2021). A giant NLR gene confers broad-spectrum resistance to Phytophthora sojae in soybean. *Nat. Commun.* 12. doi: 10.1038/s41467-021-26554-8

Wen, L., Yuan, C., Herman, T. K., and Hartman, G. L. (2017). Accessions of perennial Glycine species with resistance to multiple types of soybean cyst nematode (Heterodera glycines). . *Plant Dis.* 101, 1201–1206. doi: 10.1094/PDIS-10-16-1472-RE

Xie, M., Chung, C. Y. L., Li, M. W., Wong, F. L., Wang, X., Liu, A., et al. (2019). A reference-grade wild soybean genome. *Nat. Commun.* 10, 1–12. doi: 10.1038/s41467-019-09142-9

Xu, L., Dong, Z., Fang, L., Luo, Y., Wei, Z., Guo, H., et al. (2019). OrthoVenn2: a web server for whole-genome comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Res.* 47, W52–W58. doi: 10.1093/nar/gkz333

Yin, D., Ji, C., Song, Q., Zhang, W., Zhang, X., Zhao, K., et al. (2019). Comparison of arachis monticola with diploid and cultivated tetraploid genomes reveals asymmetric subgenome evolution and improvement of peanut. *Adv. Sci. (Weinh)* 7. doi: 10.1002/ADVS.201901672

Zheng, F., Wu, H., Zhang, R., Li, S., He, W., Wong, F. L., et al. (2016). Molecular phylogeny and dynamic evolution of disease resistance genes in the legume family. *BMC Genomics* 17, 1–13. doi: 10.1186/s12864-016-2736-9

Zhuang, Y., Wang, X., Li, X., Hu, J., Fan, L., Landis, J. B., et al. (2022). Phylogenomics of the genus *Glycine* sheds light on polyploid evolution and life-strategy transition. *Nat. Plants* 8, 233–244. doi: 10.1038/s41477-022-01102-4

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 31 July 2024
DOI 10.3389/fpls.2024.1400215

Check for updates

OPEN ACCESS

EDITED BY
Reyazul Rouf Mir,
Sher-e-Kashmir University of Agricultural Sciences and Technology, India

REVIEWED BY
Catarina Campos,
University of Evora, Portugal
Mohamed Fokar,
Texas Tech University, United States

*CORRESPONDENCE
Khaled Masmoudi
khaledmasmoudi@uaeu.ac.ae

RECEIVED 13 March 2024
ACCEPTED 13 June 2024
PUBLISHED 31 July 2024

CITATION
Ahmad M, Abdul Aziz M, Sabeem M, Kutty MS, Sivasankaran SK, Brini F, Xiao TT, Blilou I and Masmoudi K (2024) Date palm transcriptome analysis provides new insights on changes in response to high salt stress of colonized roots with the endophytic fungus *Piriformospora indica*. *Front. Plant Sci.* 15:1400215. doi: 10.3389/fpls.2024.1400215

COPYRIGHT
© 2024 Ahmad, Abdul Aziz, Sabeem, Kutty, Sivasankaran, Brini, Xiao, Blilou and Masmoudi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Date palm transcriptome analysis provides new insights on changes in response to high salt stress of colonized roots with the endophytic fungus *Piriformospora indica*

Manzoor Ahmad[1], Mughair Abdul Aziz[1], Miloofer Sabeem[1], M. Sangeeta Kutty[2], Sathesh K. Sivasankaran[3], Faical Brini[4], Ting Ting Xiao[5], Ikram Blilou[6] and Khaled Masmoudi[1]*

[1]Department of Integrative Agriculture, College of Agriculture and Veterinary Medicine, United Arab Emirates University, Al Ain, United Arab Emirates, [2]Department of Vegetable Science, College of Agriculture, Kerala Agricultural University, Vellanikkara, Thrissur, India, [3]Division of Research, Innovation, and Impact, 106B Bond Life Sciences Center, University of Missouri, Columbia, MO, United States, [4]Biotechnology and Plant Improvement Laboratory, Centre of Biotechnology of Sfax (CBS)/University of Sfax, Sfax, Tunisia, [5]College of Plant Science and Technology, Beijing Key Laboratory for Agricultural Application and New Technique, Beijing University of Agriculture, Beijing, China, [6]King Abdullah University of Science and Technology (KAUST), Biological and Environmental Sciences and Engineering (BESE), Thuwal, Saudi Arabia

Salinity is a significant threat that causes considerable yield losses in date palm. The root endophytic fungus *Piriformospora indica* has proven effective in providing salt stress tolerance to host plants. However, the underlying molecular mechanism facilitating the date palm's response to *P. indica* inoculation, and its involvement in the salt stress tolerance, remains unknown. In this study, the colonization of *P. indica* on date palm seedlings exposed to saline conditions was observed through confocal microscopy, and its impact on gene expressions was evaluated using the transcriptomic analysis. Our findings show that *P. indica* colonization reinforced the cortical cells, prevented them from plasmolysis and cell death under salinity. The RNAseq analysis produced clean reads ranging from 62,040,451 to 3,652,095 across the treatment groups, successfully assembling into 30,600 annotated genes. Out of them, the number of differentially expressed genes (DEGs) varied across the treatments: i.e., 2523, 2031, and 1936 DEGs were upregulated, while 2323, 959, and 3546 were downregulated in Salt, Fungi, and Fungi+Salt groups, respectively. Furthermore, principal component analysis based on transcriptome profiles revealed discrete clustering of samples from different treatment groups. KEGG and GO pathways enrichment analysis highlighted variation in the number and types of enriched pathways among the treatments. Our study indicated variations in gene expression related to plant hormone biosynthesis and signal transduction (auxin, abscisic acid, gibberellin, and ethylene), ABC transporters, sodium/hydrogen exchanger, cation HKT transporter, transcription factors such as WRKY and MYBs, and the plant immune system (lipoxygenase and jasmonate) of

the date palm seedlings. By characterizing the transcriptome of date palm roots under salt stress and with colonization of *P. indica*, the present findings provide valuable perspectives on the molecular mechanisms responsible for inducing salinity stress tolerance in plants.

KEYWORDS

transcriptome, *P. dactylifera*, *P. indica*, salinity, DEGs

1 Introduction

Salinity is a major challenge to agricultural crop growth, impacting approximately 20% of cultivated land and 33% of irrigated land globally (Shrivasata and Kumar, 2015). It disrupts crucial physiological and metabolic processes in plants, including photosynthesis, respiration, and protein synthesis (Athar et al., 2022). This disruption arises from ion toxicity, nutrient imbalances, decreased osmotic potential, and impaired water uptake, leading to damage to plant cell organelles (Abdul Aziz and Masmoudi, 2023a). In arid regions, increased groundwater salt levels have severely affected date palm production. Date palm, *Phoenix dactylifera*, employs intricate mechanisms involving sensing, signaling, and response to mitigate the adverse effects of salt stress (Ma et al., 2022). Therefore, enhancing date palm's defense mechanisms of salinity tolerance is crucial for ensuring efficient date palm productivity in salt-affected regions.

Certain soil fungal species contribute to the enhancement of plant salt stress defense mechanisms, due to their adaptability to saline environments (Siddiqui et al., 2022). Endophytic mycorrhizal fungi establish mutually beneficial relationships with various terrestrial plant species, offering a potential solution for overcoming salinity stress. *Piriformospora indica*, belonging to the Sebacinales order, is an endosymbiont characterized as a harmless, axenically cultivable root endophyte. This fungus has the capability to colonize the roots of numerous higher plants, improving their growth by enhancing stress tolerance and disease resistance (Jisha et al., 2019). This has led to the speculation of their potential role as a symbiotic fungus for enhancing date palm's tolerance to salinity stress.

A systematic investigation into the mechanisms underlying the date palm and *P. indica* colonization can be an essential way for improving the salt tolerance of date palm. The various pathways through which *P. indica* modulates plant physiological processes have been examined in different model interaction systems, such as *P. indica* colonization with *Arabidopsis thaliana* and *Hordeum vulgare* (Achatz et al., 2010; Qiang et al., 2012). Transcriptomic analyses in various studies have found differentially expressed genes (DEGs) in the roots of plants inoculated with *P. indica* under diverse stress conditions, including salinity stress (Ghaffari et al., 2016; Abdelaziz et al., 2019), drought stress (Zhang et al., 2018), early blight (Panda et al., 2019), and root-knot nematode infections (Atia et al., 2020). *P. indica* inoculation triggers the expression of host plant defense-related genes (*PR*, *LOX2*, and *ERF1*), abiotic stress-responsive genes (*DREB2A*, *CBL1*, and *RD29A*), and osmoprotectants like proline and glycine betaine (Waller et al., 2005; Zarea et al., 2012; Trivedi et al., 2013).

Thus, there is clear evidence that *P. indica* plays a crucial role in balancing the trade-off between the growth of host plants and their tolerance to salt stress by influencing various physiological aspects. The interaction between plants and this endophytic fungus is intricate and mutual, making it stimulating to reveal the transcriptomic alterations in date palms with the inoculation of *P. indica* in response to salt stress. Hence, it becomes imperative to develop novel molecular strategies aimed at enhancing salt stress tolerance in date palms through using endophytic fungi.

Limited transcriptome profiling studies have been conducted in non-model plants like date palm (Abdul Aziz and Masmoudi, 2024). Al-Dous et al. (2011) utilized Illumina GAII sequencing to assemble 58% of the date palm genome, predicting 25,059 genes. Bourgis et al. (2011) performed a comparative transcriptome study on oil and date palm mesocarps using pyrosequencing data from the Roche GS FLX Titanium platform. Additionally, pyrosequencing data contributed to transcriptomic profiles for date palm fruit development (Yin et al., 2012). Full-genome assemblies of date palm's plastid (158,462 bp) and mitochondrion (715,001 bp) have been achieved (Yang et al., 2010; Fang et al., 2012). Al-Mssallem et al. (2013) conducted *de novo* transcriptome assembly using second and third generation sequencing technologies, revealing stress-related genes in date palm. Despite these efforts, the transcriptomic profiling of genes in response to salt stress for *P. indica* colonized date palm remains unexplored.

In our previous study, we have demonstrated that inoculation of date palm seedlings with the beneficial endophyte *P. indica* reduced the adverse impacts of salt stress (Sabeem et al., 2022). It improved the growth of date palm seedlings through the maintenance of ion balance, enhanced uptake of nutrients, and increased antioxidant activity. Hence, this study aims to enhance our understanding of *P. indica*-mediated salt tolerance in date palms through root transcriptome analysis. The main objectives include gaining a deeper understanding of *P. indica's* potential to alleviate salt stress by profiling differentially

expressed transcripts/genes in date palm during salt stress. Through RNA sequencing, the current investigation compares the transcriptomes of *P. indica*-colonized and non-colonized date palm roots under salt stress condition. Numerous salt-responsive date palm differentially expressed genes (DEGs) were identified, and their functional annotation was conducted. This transcriptome profiling contributes to our knowledge on how *P. indica* colonization in date palm roots aids plants to resist the increasing problem of salinity stress in arid regions.

2 Materials and methods

2.1 Plant material and growth conditions

Tissue culture seedlings of the Khalas variety of date palm were obtained from the tissue culture laboratory at UAE University and grown in ½ MS liquid medium (R core team, 2022). Before exposure to salinity stress treatments, seedlings were cultivated in a growth chamber for four weeks under controlled conditions, at a temperature of 25°C ± 2°C, a photoperiod consisting of 16 hours of light and 8 hours of darkness, and a relative humidity of 60%.

2.2 Fungal growth conditions, root colonization, and stress treatments

P. indica was cultivated on potato dextrose agar (PDA) medium (Sigma, St. Louis, USA) and left to incubate for 10 days at 28°C in darkness. Chlamydospores from the agar plate were introduced to 100 ml of Kaefer medium for liquid culture in 500 ml Erlenmeyer flasks. Furthermore, the flasks were placed on a rotary shaker, operating at 100 rpm, for 8 to 10 days at 28°C. Following centrifugation of the *P. indica* liquid culture, the harvested solution had its mycelium removed and was washed three times with sterile distilled water.

For the co-cultivation process, *P. indica* was combined with the roots of *in vitro*-grown Khalas date palm seedlings. This involved injecting 1% of a 500 µl diluted mycelial solution into the media surrounding the roots, with a subsequent gentle shaking of the mixture. The inoculated seedlings were placed in a growth chamber at 25°C with a photoperiod of 16 hours of light and 8 hours of darkness. To shield the root zones from light exposure, aluminum foil covered the bottom of the tubes containing the seedlings. After allowing *P. indica* to propagate within the roots of Khalas date palm seedlings for three weeks, four experimental groups were generated. Plant seedlings grown in normal water were classified as control. Seedlings grown in ½ MS medium with 250 mM NaCl are termed as the Salt group. Those cultivated in normal water and inoculated with *P. indica* are determined as the Fungi group. Seedlings grown in a ½ MS medium with 250 mM NaCl and inoculated with *P. indica* are categorized as the Fungi+Salt group in this study. Each experimental group was prepared in triplicates and samples from each replicate was used for subsequent transcriptomic and root structure analysis.

2.3 Root structure after colonization with *P. indica* and application of salt stress

Roots were collected after three weeks from both inoculated and non-inoculated seedlings cultivated in ½ MS liquid medium with the addition of 250 mM NaCl and under control conditions. These roots were embedded in 8% low melting agarose, and cross-sections were developed with a vibratome (Xiao et al., 2019), resulting in sections with a thickness of 150 µm. Consequently, the sections were stained using the cell wall dye SCRI Renaissance (Musielak et al., 2015).

Imaging of the sections was conducted using an LSM 880 Airyscan Confocal Microscope (Leica Microsystems, Wetzlar, Germany). The 405-nm laser was employed to visualize the cell wall stain, while the 488-nm laser was used for autofluorescence. Images were captured with a 2x0.8M17 Plan-Apochromat objective, and the tile scan function was employed to obtain the full field of view.

2.4 Total RNA extraction and cDNA library preparation for RNA-seq sequencing

Total RNA extraction from root samples of different treatments was carried out using the RNeasy total RNA isolation kit from Qiagen, following the manufacturer's instructions. The concentration of RNA was determined using a NanoDrop1000 spectrophotometer from Thermo Scientific, USA. Approximately 3 µg of RNA per sample was used for cDNA synthesis. The TruSeq RNA Sample Prep Kit from Illumina, USA, was utilized to construct the cDNA libraries, following the manufacturer's protocol, and indexed codes were introduced to associate sequences with each sample. Macrogen in Seoul, South Korea, performed paired-end sequencing for all libraries using an Illumina HiSeq X-ten platform.

2.5 Quality control and data analysis

The raw data in fastq file format were examined with the quality filtering procedures. Using the fastp tool, adaptor sequences and low-quality reads (with a quality score below Q20) were eliminated to obtain clean reads. These clean reads were used in subsequent analyses for downstream steps. For each sample, high-quality clean reads were aligned to the reference *P. dactylifera* genome (https://www.ncbi.nlm.nih.gov/bioproject/322046) using the HISAT2 tool. The feature counts tool was used to quantify transcript abundance and read counts (Liao et al., 2014). Moreover, gene expression levels were assessed through calculations using FPKM (fragments per kilobase of transcript sequence per million base pairs), a metric that normalizes expression by considering the transcript length and total mapped reads.

2.6 Differential expression analysis and functional annotation of transcripts

Differential expression analysis was performed using the DESeq2 R package (version 1.43.1) (Love et al., 2014). DEGs were identified by comparing the expression levels of all transcripts between the salt stressed and control groups. Genes exhibiting a fold change greater than 1 and a false discovery rate (FDR) value below 0.05 were considered differentially expressed.

The functions of the identified DEGs were annotated using the Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), and Uniprot databases. Functional gene set enrichment analysis was conducted using the iDEP.96 software available at sdstate.edu. This analysis provided perceptions into the enriched biological processes and pathways associated with the identified DEGs, aiding in the interpretation of their functional significance in the context of the experimental conditions.

2.7 Data processing and visualization

The data analysis and visualization were carried out using R (version 4.3.1). Various packages were used for different aspects of analysis and visualization (R core team, 2022). Specifically, the Pheatmap package was utilized for generating heatmaps, the ggplot2 package for creating bar plots and volcano plots, and the ggVennDiagram package for generating Venn diagrams. The use of diverse R packages allowed for a thorough exploration and presentation of the results obtained from the differential expression analysis and functional annotation of genes.

3 Results

3.1 Root structural characterization of the date palm seedlings

To understand the cellular and subcellular responses of *P. indica* colonization in living date palm roots challenged with 250 mM NaCl, confocal microscopy was used to visualize tissue anatomy and reveal the consequences of salt treatment and the colonization of roots with *P. indica*. The *P. indica* colonized cortical cells reinforced the root architecture and prevented the plasmolysis and cell death of salt treated roots (Figure 1). The non-colonized and salt treated roots showed huge plasmolysis of the cortex cell layers, whereas the colonized roots with *P. indica* and salt-treated preserved the cortex cell layers from plasmolysis (Figure 1).

3.2 RNA sequencing analysis of the date palm seedlings root

A transcriptomic analysis was conducted to investigate the molecular mechanisms underlying *P. indica*-mediated salt tolerance or adaptation in date palm seedlings under salinity stress. The sequencing results yielded raw reads ranging from 67,295,188 to 39,977,852 (Supplementary Table S1). After the elimination of low-quality reads and contaminants, the clean reads from all samples ranged from 62,040,451 to 3,652,095 (Supplementary Table S1). The control group had the highest number of both raw and clean sequencing reads, while the Fungi+Salt treatment group had the lowest (Supplementary Table S1). Similarly, the control group exhibited the highest number of mapped reads at 5,254,737, whereas the Fungi+Salt group had the lowest reads (3,275,435). Approximately, 30,600 functional genes were annotated from the clean reads, with over 94% successfully assigned to protein-coding genes across all samples (Figure 2). Particularly, the Fungi+Salt group showed a relatively higher number of protein-coding genes (94% to 96%) compared to other treatments.

3.3 Identification and evaluation of DEGs

In order to understand how *P. indica* colonization enhances salt tolerance in date palms, we analyzed the DEGs in the established experimental groups. The transcriptomic profile of the control group (without salt and uncolonized) was used as a reference for both salt and fungi treatments. However, for the Fungi+Salt group, the salt treatment was used as a reference transcriptome. DEGs were examined based on a fold change of ≥ 1 and an FDR of ≤ 0.05 in response to salt stress and

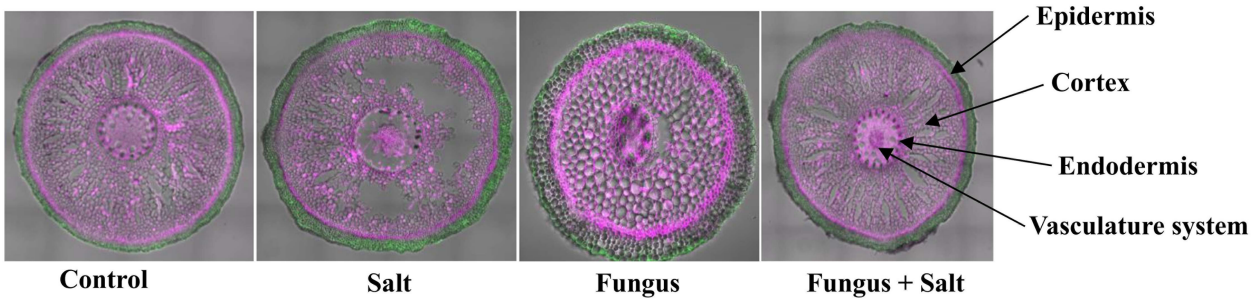


FIGURE 1
Anatomical structure of date seedlings roots. Confocal microscopy was used to examine the tissue anatomy of date palm roots colonized and non colonized with *P. indica* under control and salt stress conditions. The cell wall was stained with a pink color, while autofluorescence was represented by a green color.

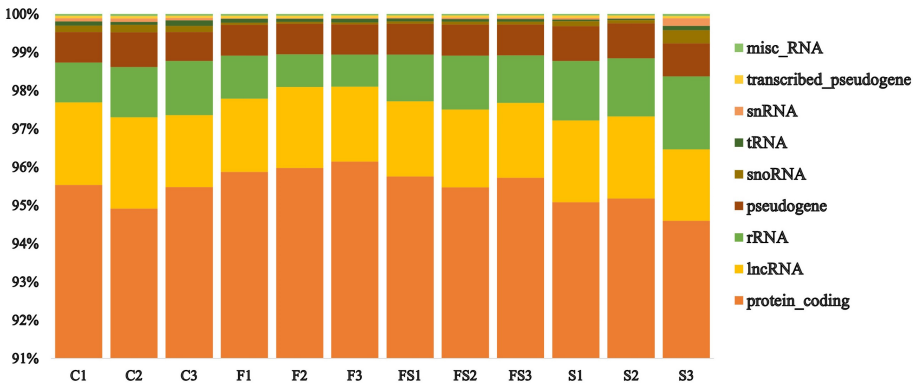


FIGURE 2
Distribution of the functionally assigned clean reads of the date palm seedlings root RNA Sequencing. The functionally assigned clean reads were distributed across various categories, including protein coding, long non-coding RNA (lncRNA), ribosomal RNA (rRNA), pseudogene, small nucleolar RNA (snoRNA), transfer RNA (tRNA), small nuclear RNA (snRNA), transcribed pseudogene, and miscellaneous RNA (misc_RNA), in different treatments. The treatments were categorized as follows: C: control), F: *P. indica* colonized date palm seedlings under control condition, FS: *P. indica* colonized date palms under salt stress condition, and S: non-colonized date palm seedlings under salt stress conditions.

P. indica colonization in date palm roots. In the Salt, Fungi, and Fungi +Salt treatments, a total of 25,902, 23,351, and 25,679 DEGs were identified, respectively (Figures 3A–C; Supplementary Table S2).

Among the treatments, a high number of upregulated DEGs were observed in the Salt group (2523) followed by fungi group (2301), whereas the lowest number (1936) was observed in Fungi+Salt group (Supplementary Table S2). Conversely, Fungi+Salt treatment displayed the highest count of downregulated DEGs (3546), trailed by the Salt group (2323), whereas only 959 downregulated DEGs were identified in the fungi treated group. The variation in DEGs between the treatments showed how the colonization of *P. indica* influenced the date palm's functional responses under normal condition and salinity stress. Moreover, it is worth noting that, among the upregulated DEGs, only 19 genes (0.4%) were observed to be present in all treatment groups (Figure 4A). In contrast, most upregulated DEGs exhibited treatment-specific expression patterns, as depicted in the Venn diagram (Figure 4A). Correspondingly, when considering the downregulated DEGs, it is notable that just 334 genes (6.8%) were detected consistently across all treatments. Moreover, the salt group and Fungi+salt group exhibit the largest proportion of upregulated and downregulated DEGs, with 1020 (23.0%) and 944 (19.3%) DEGs respectively (Figures 4A, B). The PCA analysis based on the transcriptome profile showed the distinct clustering of the samples of different treatments, which further explained the differences in functional responses of date palm under different stress conditions (Figure 4C).

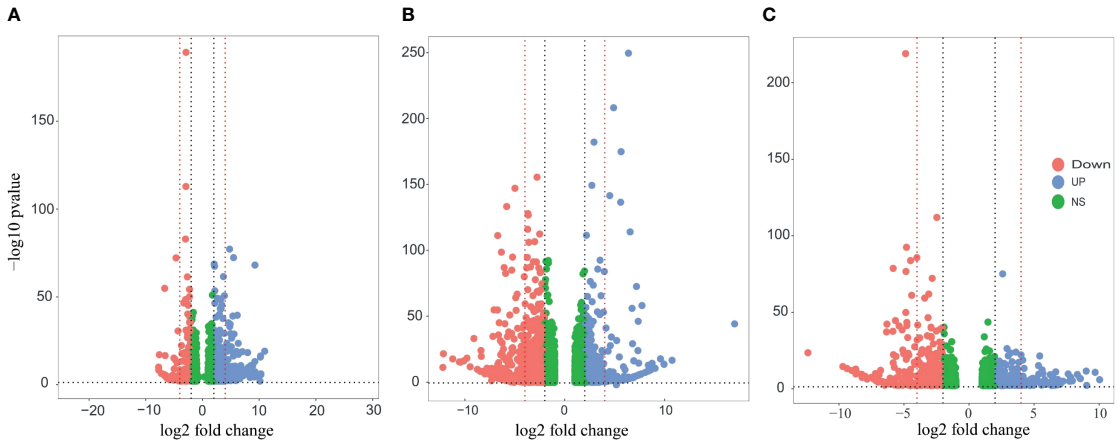


FIGURE 3
Volcano plots of the DEGs in the different treatment groups of the date palm seedlings. DEGs were identified using DESeq2 R package (version 1.43.1). The vertical axis represents the statistical significance of the difference in gene expression. Blue dots indicate the upregulated DEGs, the red dot signifies downregulated DEGs, and the green dot shows the non-significant DEGs. **(A)** Visualization of the DEGS in the date palm seedlings inoculated with *P. indica* under control condition. **(B)** Visualization of the DEGS in the non-inoculated date palm seedlings inoculated under salt stress condition. **(C)** Visualization of the DEGS in the date palm seedlings inoculated with *P. indica* under salt stress condition. The analysis revealed significant variation in the expression pattern of DEGs across the treatments.

10.3389/fpls.2024.1400215

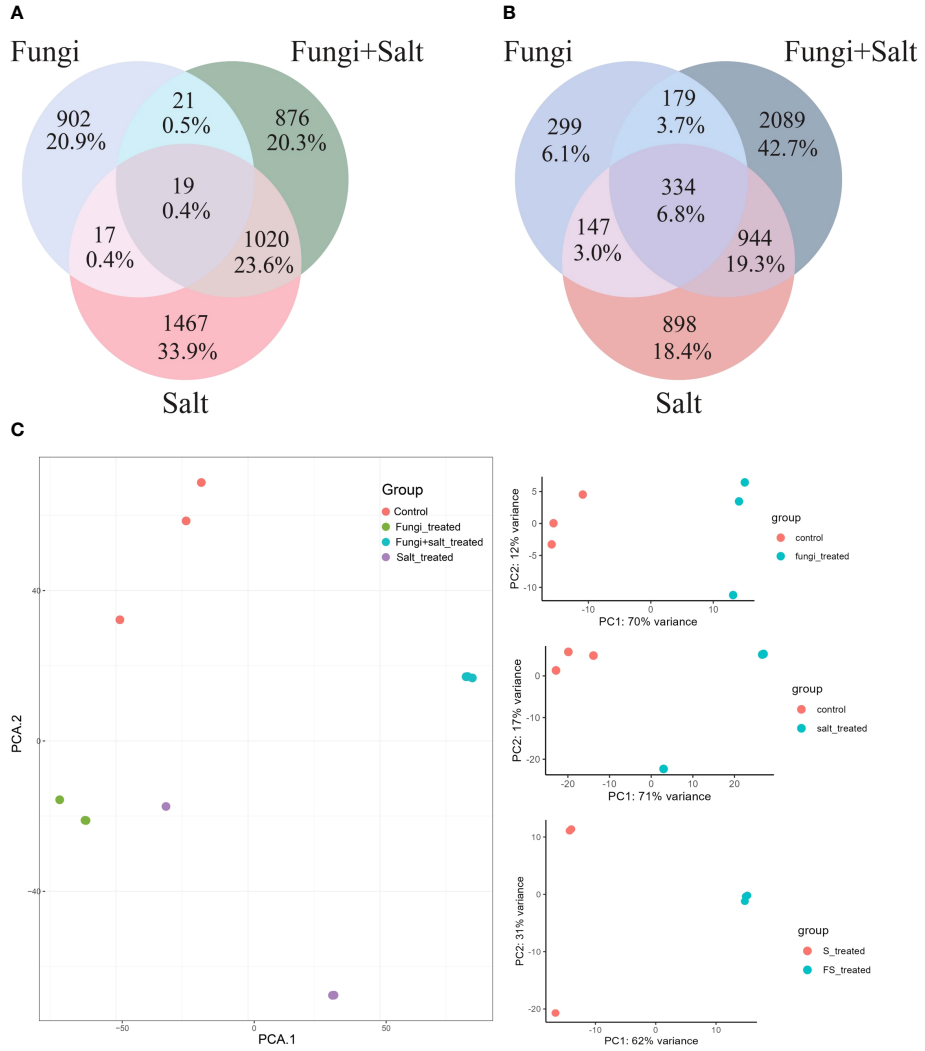


FIGURE 4
Variances and similarities in the transcriptomic profile of DEGs across the treatment groups. DEGs were identified using DESeq2 R package (version 1.43.1). In this analysis only DEGs with fold change ≥ 1 and ≤ -1 are included. **(A)** Venn diagram displaying the shared and treatment specific upregulated DEGs among the treatments. **(B)** Venn diagram showing the shared and treatment-specific downregulated DEGs across the treatments. **(C)** PCA based on the Fragments Per Kilobase Million (FPKM) values demonstrating the clustering of samples from the different experimental groups.

3.4 Pathway enrichment analysis of the transcriptome of the colonized date palm with *P. indica*

Various source databases, such as NCBI, GO, KEGG, and UniProt, were used for pathway enrichment analysis and functional gene annotation. The analysis results indicate variations in gene counts and fold changes associated with different pathways in the established treatments (Figures 5A–C).

For the pathway enrichment analysis and functional gene annotation, different source databases were used, including the NCBI, GO, KEGG and UniProt. Overall, the analysis findings show variations in gene counts and fold changes of genes associated with different pathways in the established treatments. For example, we observed a higher number of enriched KEGG pathways in the Salt+Fungi group (18), followed by the Fungi group (17) (Figure 5A), whereas the lowest number (16) enriched KEGG pathways were observed in the Salt group. Additionally, the different groups exhibited variations in the enriched pathways. For instance, the Fungi group showed significant enrichment in pathways, including DNA replication (3.86 folds), phagosome (1.75 folds), cysteine, and methionine metabolism (1.65 folds), stilbenoid, diarylheptanoid and gingerol biosynthesis (3.70 folds), starch and sucrose metabolism (2.14 folds), as well as biosynthesis of various plant secondary metabolites (3.76 folds) (Figure 5B). Some pathways, including fatty acid elongation (2.91 folds), fatty acid degradation (2.00 folds), diterpenoid biosynthesis (2.46 folds), steroid biosynthesis (1.99 folds), alpha-Linolenic acid metabolism (1.85 folds), and brassinosteroid biosynthesis (2.97 folds), were enriched in the Fungi+Salt treated group (Figure 5C). Additionally, we found that the pathways associated with ABC transporters were present in both the Fungi+Salt and Fungi groups but were not detected in the Salt group. The GO pathway enrichment analysis revealed a higher number of enriched

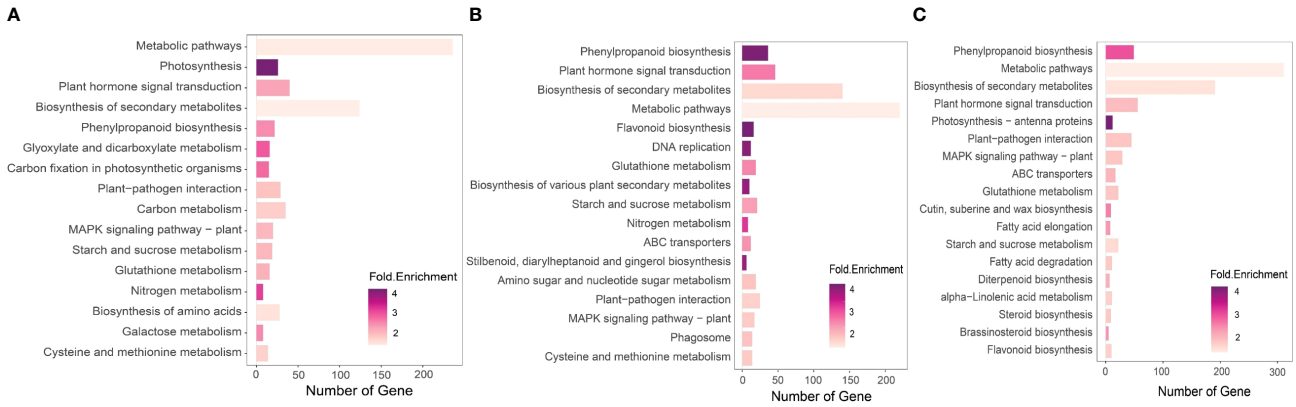


FIGURE 5
KEGG pathway enrichment analysis of the DEGs of the experimental date palm groups. We performed KEGG pathway enrichment analysis using iDEP.96 available at iDEP.96 (sdstate.edu). The analysis revealed variation in the number and types of enriched pathways in the experimental groups. **(A)** Non-inoculated date palm seedlings under salt stress (Salt group) **(B)** *P. indica* inoculated date palm seedlings under salt stress (Fungi) **(C)** *P. inidca* inoculated date palm seedling under the control condition (Fungi+Salt).

pathways in the Fungi+Salt group (Figure 6A), followed by the Salt group (Figure 6B). The Fungi treated group exhibited a lower number of enriched pathways (3) compared to Fungi+Salt (14) and Salt (8) groups (Figure 6C). In a similar way, the Uniprot pathway enrichment analysis revealed a clear and discernible pattern of enriched pathways across the various experimental treatments. The discrepancy in enriched pathways demonstrates the impact of *P. indica* colonization on the functional responses of date palms under salinity stress (Supplementary Figure S1).

3.5 Impact of *P. indica* inoculation on the date palm hormone related genes

Plant hormones regulation is important for plant cell cycle, plant growth and stress responses. The analysis of DEGs related to plant hormones, including auxin, gibberellin, and abscisic acid, was conducted in the transcriptome of each experimental group. The results revealed specific variations in the expression profiles of genes associated with hormone regulation across the treatments. Particularly,

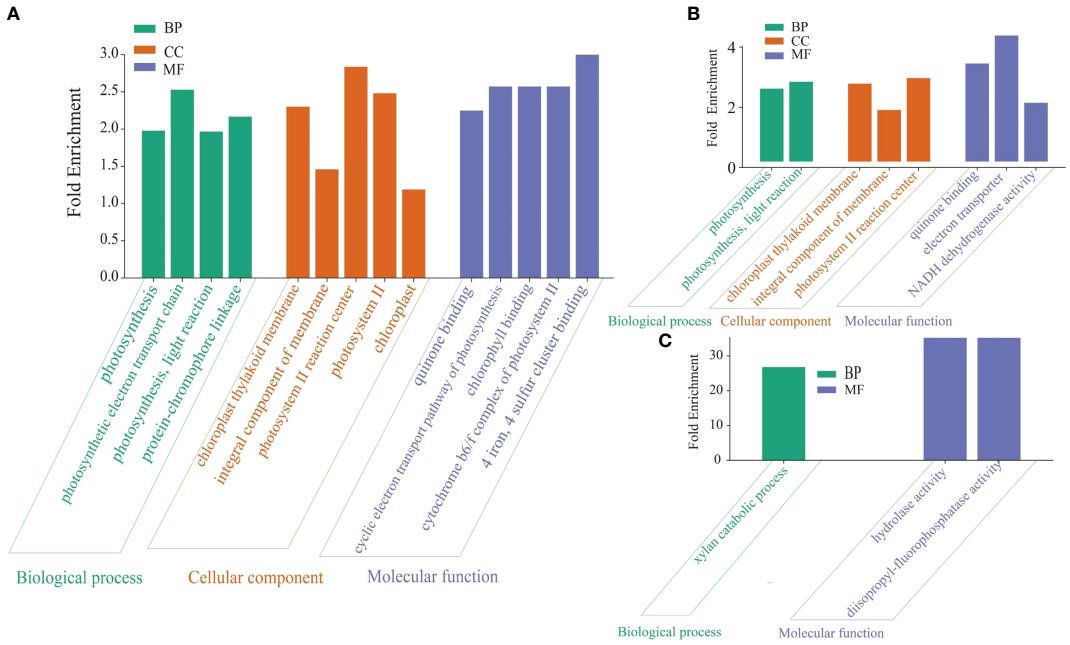


FIGURE 6
GO pathway of the DEGs in the date palm seedlings treatment conditions. The GO pathway enrichment analysis was performed using iDEP.96 tool available at iDEP.96 (sdstate.edu). The analysis showed substantial variations in enriched pathways across the experimental groups. **(A)** *P. indica* inoculated date palm seedlings under salt stress condition (Fungi+Salt group). **(B)** Non-inoculated date palm seedlings under salt stress condition (Salt group). **(C)** *P. indica* inoculated seedlings under the control condition (Fungi group).

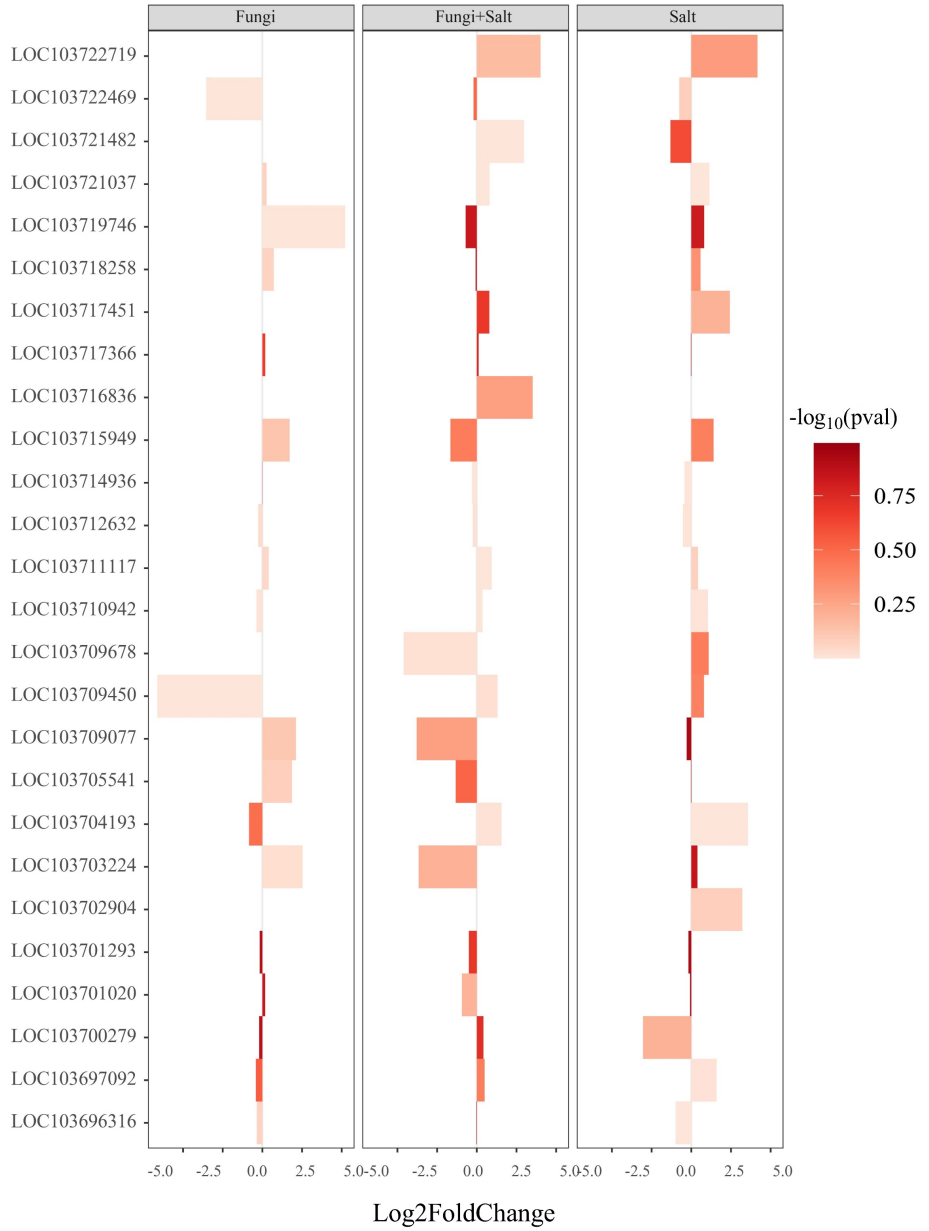


FIGURE 7
Bar plot showing the expression of DEGs associated with gibberellin biosynthesis and catabolism. DEGs were identified using DESeq2 R package (version 1.43.1). DEGs involved in gibberellin biosynthesis and catabolism were retrieved from the DEGs profile of each treatment and a heated bar plot were generated using ggplot2 package in R. The coordinates of the figure show the statistics of genes expression. The bar color intensity shows the increase and decrease of *P* value in the form of -$\log_{10}$. Fungi: *P. indica* inoculated under control condition, Fungi+salt: *P. indica* inoculated under salt stress condition, and Salt: Non- inoculated salt stress condition.

a larger number of DEGs related to gibberellin biosynthesis were identified in the *P. indica* inoculated date palm seedlings under the salt stress condition (28), followed by the non-inoculated seedlings under the salt stress condition (25), while the *P. indica* inoculation under the control condition exhibited the lowest number (18) (Figure 7; Supplementary Table S3). Particularly, genes associated with gibberellin 2-beta-dioxygenase 8 (1.10 folds), gibberellin 2-beta-dioxygenase 2 (3.64 folds), and gibberellin 3-beta-dioxygenase 1-like (2.47 folds) were significantly upregulated in the *P. indica* inoculated date palms under salt stress treatment. Moreover, genes including gibberellin 2-beta-dioxygenase 1 (1.74 folds), gibberellin 20-oxidase-like protein (1.86 folds), gibberellin-regulated protein 5 (2.11 folds), and gibberellin-regulated protein 6 (5.46 folds) were significantly upregulated in the *P. indica* inoculated date palm seedling under control condition but downregulated in the *P. indica* inoculated date palm seedling under salt stress condition (Figure 7).

Similarly, significant differences were observed in the DEGs associated with auxin and abscisic acid-responsive proteins across the treatments (Supplementary Figures S2, S3). The *P. inidca* inoculated date palm seedlings under the salt stress condition showed a higher abundance of DEGs related to auxin (75) and abscisic acid (78), followed by the non-inoculated salt stress

10.3389/fpls.2024.1400215

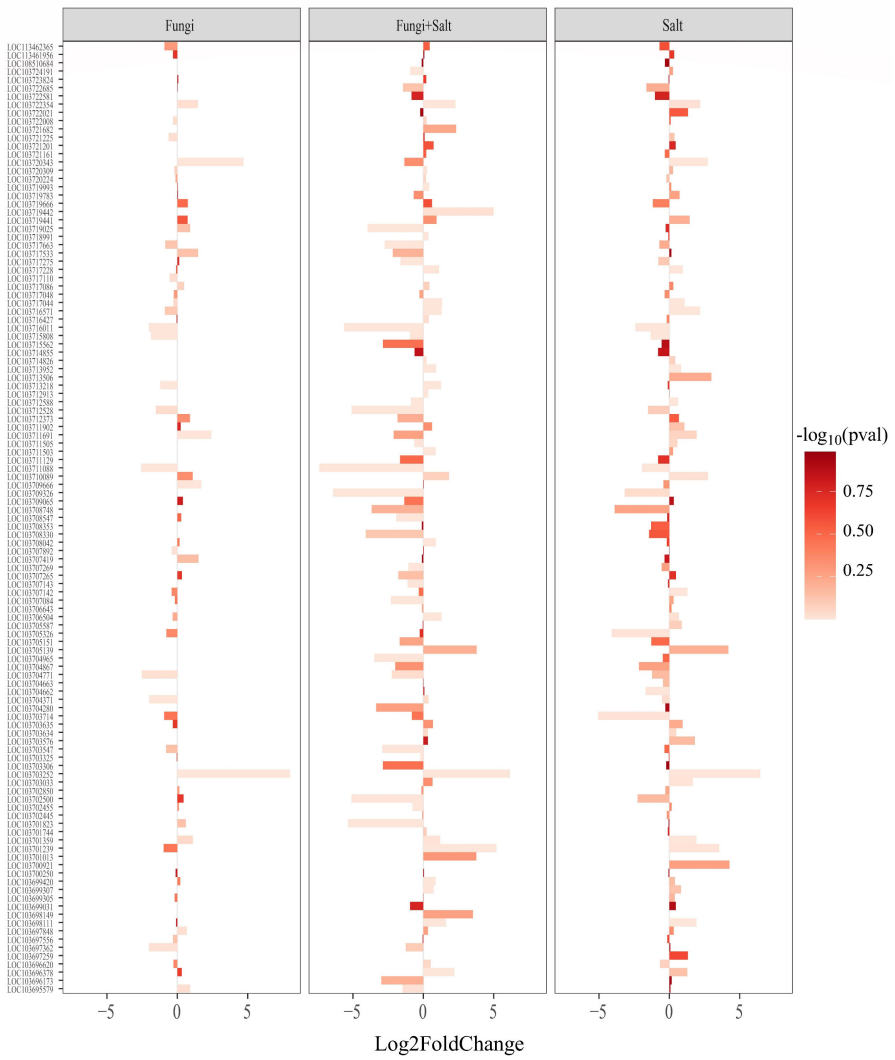


FIGURE 8
Bar plot showing the expression pattern of ethylene-responsive transcription factors (ERFs) across the treatments. DEGs were identified using DESeq2 R package (version 1.43.1). DEGs related to ERFs were retrieved from the DEGs profile of each treatment and a heated bar plot were generated using ggplot2 package in R. The bar coordinates show the statistics of ERFs expression. The bar color intensity shows the increase and decrease of *P* value in the form of -$\log_{10}$. Fungi: *P. indica* inoculated under control condition, Fungi+salt: *P. indica* inoculated under salt stress condition, and Salt: Non- inoculated salt stress condition.

condition (79 and 77), and the *P. indica* inoculated under the control condition (57 and 44) (Supplementary Figures S2, S3; Supplementary Tables S4, S5). These findings highlight the intricate and treatment-specific regulation of plant hormone-related genes, emphasizing their importance in plant cell cycle, growth, and stress responses.

3.6 *P. indica* inoculation alters the expression pattern of transcription factors

In this study, stress tolerance was explored with a focus on transcription factors, crucial regulatory components in stress-responsive pathways. Specifically, three families of stress-responsive transcription factors—ethylene-responsive transcription factors (ERFs), WRKYs, and MYBs—were examined in the transcriptome of the treatment groups. The results revealed significant variations in the number and expression patterns of DEGs associated with these transcription factor families across the treatments (Figure 8).

Among the three families, the *P. indica* inoculated date palm seedlings under the salt stress condition showed the highest number of ERFs (111), followed closely by the non-inoculated seedlings under the salt stress condition with 110 ERFs, while the *P. indica* inoculated date palm seedling under the control condition showed 68 ERFs (Figure 8; Supplementary Table S6). Furthermore, the expression of ERFs within each treatment varied, for instance, in the *P. indica* inoculated date palm seedlings under the salt stress, 17 out of 111 ERFs (15.3%) were upregulated, and 34 (30.6%) were downregulated. Similarly, in the non-inoculated salt stress condition and inoculated

control condition, 21 (19.1%) and 9 (13.2%) ERFs were upregulated, while 18 (16.3%) and 8 (11.7%) were downregulated, respectively (Supplementary Table S6).

In the other two transcription factor families (WRKYs and MYBs), significant variations were observed in both the number and expression patterns across the experimental treatments. Particularly, the *P. indica* inoculated under the salt stress group had the highest number of WRKYs (88) and MYBs (34), followed by the non-inoculated salt stress group with 85 WRKYs and 33 MYBs, and the *P. indica* inoculated under the control condition with 71 WRKYs and 30 MYBs (Supplementary Figures S4, S5; Supplementary Tables S7, S8). These findings indicate the intricate regulatory role of transcription factors in stress responses and highlight treatment-specific differences in their expression profiles.

3.7 *P. indica* inoculation modulates the expression of immune and transporter genes

The inherent immune system of plants relies on various components, including salicylic acid, jasmonate, and ethylene, which play pivotal roles in influencing both harmful pathogens and symbiotic organisms. Additionally, lipoxygenases contribute by converting linoleic or alpha-linolenic acid into (9S)- and (13S)-hydroperoxides, serving as precursors for various oxylipins involved in plant defense. Transcriptomic analysis indicated that the colonization of *P. indica* led to alterations in the abundance and expression of genes related to plant defense.

The expression of genes involved in jasmonate synthesis exhibited differences among the experimental groups (Supplementary Figure S6; Supplementary Table S9). The non-inoculated salt stress group showed the highest number of DEGs related to jasmonate (4), followed by *P. indica* inoculated under the salt stress condition (3), and *P. indica* inoculated under the control condition (2). Similarly, a total of 8 DEGs related to lipoxygenases were identified in the transcriptome of the experimental groups. Specifically, a significant variation was observed in the expression patterns of these DEGs across the treatments (Supplementary Figure S6; Supplementary Table S10). In the *P. indica* inoculated seedlings under the salt stress condition, 6 out of the jasmonate-related DEGs (75%) were downregulated, while in the non-inoculated salt stress condition, only 3 (37.5%) were downregulated.

Transporter genes, including ABC transporters, HKT, and sodium/hydrogen exchangers, play a crucial role as transporter proteins in plant stress responses and various physiological processes (Sabeem et al., 2022). In this study, three differentially expressed genes (DEGs) related to the HKT gene, specifically HKT1, HKT6, and HKT8, along with their isoforms, were identified in the transcriptome of treatment condition (Supplementary Figure S7; Supplementary Table S11). These genes exhibited a differential expression pattern across the treatments. Moreover, all HKT-related genes were upregulated in the *P. indica* inoculation under control condition, while they were downregulated in the *P. indica* inoculated under the salt stress condition.

Similarly, four DEGs related to sodium/hydrogen exchangers and their isoforms were identified with varying expression levels in different treatments (Supplementary Figure S8; Supplementary Table S12). In the *P. indica* inoculation under the control condition, most of the sodium/hydrogen exchangers showed downregulation. However, a different expression pattern was observed in the *P. indica* inoculation under the salt stress and non-inoculated salt stress condition, except for one isoform (LOC103713094: sodium/hydrogen exchanger 2-like), which exhibited upregulation in both groups, (1.28 folds) and (1.11 folds), respectively.

ABC transporters play an essential role in plant physiology by facilitating the movement of various substances across cell membranes. This involvement is critical for essential processes such as nutrient absorption, hormone signaling, detoxification, and plant-microbe interactions, among others. Ultimately, these transporters significantly impact plant growth, development, and stress responses. In the present study, 109, 107, and 97 DEGs related to ABC transporters genes and their isoforms were identified in the Salt group, Fungi+Salt group and Fungi group, respectively (Supplementary Figure S8; Supplementary Table S13). The Salt group shared the highest proportion of upregulation (23.8%) followed by Salt+Fungi (21.4%) and Fungi (17.5%). On the other side, Fungi+Salt group exhibited the largest proportion of downregulated DEGs (14.5%), followed by Salt group (11%) and Fungi group (5.1%). In addition, we observed differences in the ABC transporters linked to DEGs among the treatments (Supplementary Table S13).

4 Discussion

The endophytic fungus, *P. indica*, plays a crucial role in effectively modulating plant responses to salt stress and reducing its adverse effects on plant functioning. Transcriptomic analysis serves as a gold standard proxy for exploring stress-responsive genes and understanding the molecular mechanisms of stress tolerance in plants (Abdul Aziz et al., 2022). In this study, the roots of date palm seedlings were inoculated with *P. indica*, and RNAseq was conducted after successful colonization, with the goal of revealing the molecular mechanisms underlying *P. indica*-mediated salt tolerance in date palms. The study showed the formation of a symbiotic interaction between the *P. indica* and date palm roots due to its ability to colonize a broad range of host, previously found with barley (Baltruschat et al., 2008), and rice (Jogawat et al., 2013).

An initial assessment of root sections using confocal microscopy displayed significant plasmolysis of the cortex cell layers in non-colonized salt-treated roots. In contrast, the roots colonized by *P. indica* exposed to salt treatment exhibited reinforced and preserved cortex cell layers, preventing plasmolysis. This observation displays the colonization and interaction of *P. indica* with the date palm roots, suggesting a potential protective role of *P. indica* in maintaining the structural integrity of root cells under salt stress conditions (Gill et al., 2016). Likewise, it was reported by Ali et al. (2022) that the endophytic

fungus *Stemphylium lycopresici* inoculated to maize roots under saline condition (Gill et al., 2016), enhanced the root hair development, lowered the formation of lysogenic aerenchyma and decreased the Na^+/K^+ ratio in maize plants.

The results of RNAseq showed alterations in gene expression within the roots of date palm CV Khalas colonized by *P. indica*, enhancing its tolerance to salt stress. The observed changes in gene expression patterns between *P. indica*-colonized and non-colonized date palm seedling roots under salt stress highlight the significant influence of *P. indica* colonization on plant gene expression. These findings associated with existing studies on the impact of endophytic microorganisms on host plant gene expression under abiotic stresses (Masmoudi et al., 2021). For instance, Cheng et al. (2021) reported significant gene alterations in the leaves of *Achnatherum inebrians* under salt stress induced by the fungus *Epichloe gansuensis*. In contrast, the introduction of *Epichloe coenophiala* into tall fescue resulted in minimal genetic alterations, with the DEGs primarily involved in the plant's response to abiotic stress (Dinkins et al., 2017). Similarly, the analysis of the whole transcriptome of rice plants revealed that approximately 0.5% of all genes experienced expression level changes following inoculation with *Glomus intraradices* (Güimil et al., 2005). Gazara et al. (2020) reported *P. indica*-induced gene modifications in rice plants exposed to salinity. Furthermore, Rho et al. (2018) conducted a meta-analysis, indicating that incorporating endophytes, particularly *P. indica*, positively impacted plant growth and the ability to withstand abiotic stresses such as salinity, drought, and nitrogen deficiency. These collective findings indicated the potential of *P. indica* in influencing the molecular responses of plants to environmental stresses, including salt stress, thereby enhancing their tolerance. Similarly, our results showed the salt stress tolerance was conferred in date palm by *P. indica* inoculation via the expression of defense-related genes.

The pathway enrichment analysis conducted in this study, with a specific focus on KEGG pathways, revealed significant differences in enriched pathways when comparing the transcriptomes of *P. indica*-colonized roots to non-colonized roots (salt stress) of date palm seedlings. These differences highlight the selective involvement of the *P. indica* influenced expressed in different KEGG pathways, contributing to increased resilience to salinity stress. For example, there was a significant upregulation of genes involved in phenylpropanoid biosynthesis in the *P. indica* inoculated date palm seedlings under the salt stress group (fold change = 3.95) and inoculated control condition (fold change = 3.97), compared to the non-inoculated salt stress treatment (fold change = 2.48). This indicates the beneficial role of *P. indica* colonization in elevating salt stress. Phenylpropanoids, synthesized from phenylalanine in plants, are secondary metabolites crucial for enhancing plant's ability to withstand various biotic and abiotic stresses (Korkina, 2007). Additionally, the upregulation of genes associated with flavonoid biosynthesis in *P. indica*-colonized roots (control and salt stress conditions) suggests a potential enhancement of the antioxidant capacity of date palms against reactive oxygen species (ROS) produced during salt stress. Flavonoids, known as low molecular weight phenolic compounds, possess the ability to scavenge ROS generated in plants facing biotic or abiotic stress conditions (Shomali et al., 2022).

The transcriptome analysis of date palm roots during the interaction with *P. indica* revealed a stimulation of phytohormones such as gibberellin, auxin, and abscisic acid biosynthesis and signal transduction. The expression of genes involved in gibberellin biosynthesis displayed significant differences between the roots of *P. indica*-colonized and non-colonized plants. The higher expression of gibberellin 2-beta-dioxygenase 8, gibberellin 3-beta-dioxygenase 1, and gibberellin 2-beta dioxygenase 2 in the roots of *P. indica*-colonized date palm seedlings suggests the beneficial effect of *P. indica* on plant growth under salt stress. Particularly, the presence of gibberellin 2-beta-dioxygenase 8, an enzyme involved in the hydroxylation of gibberellin precursors, exclusively in the *P. indica* inoculation under the salt stress group is essential. The role of gibberellin in interactions of endophytes with roots is well known. The earlier studies on *P. indica* mainly focused on the activation of innate immune responses through gibberellin synthesis in plants. It has been displayed that *P. indica* supported the rice plants to survive root herbivory, and gibberellin functioned as a signal and interaction component for the inducible plant tolerance against the stress conditions (Cosme et al., 2016). In barley mutants, the impaired synthesis of gibberellins affected the perception and reduced *P. indica* colonization which indicated that gibberellin acts as a modulator of the basal defense of the roots (Schäfer et al., 2009). Previous research has further showed that plants with elevated levels of gibberellin catabolism genes can better withstand challenging environments such as drought and high salinity (Zhou and Underhill, 2016).

Similarly, auxin and abscisic acid are vital plant hormones that play fundamental roles in plant metabolism, growth, and development (Bouzroud et al., 2019). The results indicate that *P. indica* colonization alters the gene expression of auxin and abscisic acid-related genes. Auxins such as IAA are not only involved in key developmental processes but also plays a role in stress defense responses (Naser and Shani, 2016). Therefore, it was not surprising in our study that root-interaction of *P. indica* with date palm seedlings interfered with its auxin metabolism or signaling, which stimulated their growth and developmental processes. Abscisic acid-intensive 5 protein, which participates in drought stress response, showed a significant upregulation (fold change = 6.43) in the *P. indica*-colonized salt-treated date palm roots. In general, ABA stimulates Arbuscular Mycorrhiza (AM) symbiosis but the impact of this hormone is largely dependent on the developmental stage of the interaction and the conditions of stress (López-Ráez, 2016). Mainly, ABA plays its role when plants and their interacting symbionts are under the stress conditions, specifically osmotic stress. Therefore, the variations in the gene expression profile related to auxins and ABA signify the beneficial effect of *P. indica* colonization on date palm growth under salt stress. The modulation of hormone-related genes by *P. indica* suggests a potential mechanism through which the fungus enhances the salt tolerance of date palm seedlings.

Transcription factors, including ERFs, WRKYs, and MYBs, exert a significant impact on regulating the biosynthesis and

signaling of stress-related hormones such as ethylene, abscisic acid, jasmonate, and salicylic acid (Abdul Aziz and Masmoudi, 2023b; Zhou et al., 2022). It has been reported that inoculation of *P. indica* with barley seedlings upregulated specific genes such as WRKY transcription factors (Molitor et al., 2011). This suggests WRKY a significant interaction factor in the *P. indica's* inoculation with date palm seedlings. Our data analysis revealed a significant difference in the expression pattern of transcription factors between *P. indica*-colonized and non-colonized roots of date palm seedlings. These variations in transcription factor expression could potentially influence the expression of key genes involved in the biosynthesis and signaling of stress-related hormones that may play a role in the salinity tolerance of date palm seedlings (Abdul Aziz et al., 2021).

For example, ethylene, a crucial phytohormone with vital roles in plant physiological processes, including growth, development, and stress responses, exhibited higher expression of 1-aminocyclopropane-1-carboxylate oxidase 1 in the *P. indica*-colonized groups (control and salt stress) compared to non-colonized roots (salt stress) (Supplementary Figure S9). 1-aminocyclopropane-1-carboxylate oxidase 1 is an essential enzyme in ethylene biosynthesis and serves as a rate-limiting factor. Therefore, our data here showed that *P. indica* induces ethylene synthesis in date roots after inoculation, which suggests that ethylene signaling is required for symbiotic interaction between them. It has been found that impaired ethylene signaling resulted in reduced root colonization in Arabidopsis (Khatabi et al., 2012). Furthermore, previous research indicated that the colonization of barley roots by *P. indica* led to the upregulation of 1-aminocyclopropane-1-carboxylate oxidase (Schäfer et al., 2009). Similarly, genes involved in jasmonate biosynthesis, a major component of the plant innate immune system, exhibited altered expressions in *P. indica*-colonized roots compared to non-colonized date palm roots. These findings are related to previous studies, suggesting that *P. indica* colonization leads to the downregulation of genes involved in jasmonate biosynthesis, resulting in a lower immune response (Schäfer et al., 2009). This downregulation facilitates successful colonization and the establishment of a symbiotic association.

5 Conclusion

This study provides evidence that the colonization of *P. indica*, a fungal endophyte, in date palm roots contributes to the reinforcement of root architecture, prevention of plasmolysis, and reduction in cell death under high salinity conditions. Transcriptome analysis of the roots revealed that *P. indica* colonization led to the upregulation of numerous genes associated with salt stress and signaling pathways. Results from the GO and KEGG analyses displayed DEGs associated with various biological processes crucial for responding to salinity tolerance. Furthermore, the modulation of gene expression related to gibberellin, auxin, ethylene biosynthesis, and signaling, and with varied expression patterns of downstream transcription factors, is anticipated to exhibit a strong correlation with salt stress tolerance. The findings of this study provide essential insights into the molecular mechanisms underlying *P. indica*-mediated salt tolerance in date palm. It lays the groundwork for future research on the interaction between date palm and *P. indica*, as well as other endophytic microorganisms, an area of study that has received limited attention.

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

MA: Formal analysis, Software, Visualization, Writing – original draft, Writing – review & editing. MAA: Investigation, Methodology, Writing – review & editing. MS: Methodology, Writing – review & editing. MSK: Investigation, Methodology, Writing – review & editing. SKS: Formal analysis, Writing – review & editing. FB: Investigation, Writing – review & editing. TX: Methodology, Writing – review & editing. IB: Investigation, Methodology, Writing – review & editing. KM: Conceptualization, Funding acquisition, Resources, Supervision, Validation, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research work was supported by funding from the United Arab Emirates University Research Office grants (#31R203 and #12F041) to KM.

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2024.1400215/full#supplementary-material

References

Abdelaziz, M. E., Abdelsattar, M., Abdeldaym, E. A., Atia, M. A., Mahmoud, A. W. M., Saad, M. M., et al. (2019). *Piriformospora indica* alters Na?/K? homeostasis, antioxidant enzymes and LeNHX1 expression of greenhouse tomato grown under salt stress. *Sci. Hortic.* 256, 108532f. doi: 10.1016/j.scienta.2019.05.059

Abdul Aziz, M., Brini, F., Rouached, H., and Masmoudi, K. (2022). Genetically engineered crops for sustainably enhanced food production systems. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.1027828

Abdul Aziz, M., and Masmoudi, K. (2023a). Insights into the transcriptomics of crop wild relatives to unravel the salinity stress adaptive mechanisms. *Int. J. Mol. Sci.* 24, 9813. doi: 10.3390/ijms24129813

Abdul Aziz, M., and Masmoudi, K. (2023b). "Multifaceted roles of versatile LEA-II proteins in plants," in *Multiple abiotic stress tolerances in higher plants*. Eds. N. K. Gupta., Y. Shavrukov., R. K. Singhal and N. Borisjuk (CRC Press, Boca Raton, FL), 143–161. doi: 10.1201/9781003300564–10

Abdul Aziz, M., and Masmoudi, K. (2024). Molecular breakthroughs in modern plant breeding techniques. *Hortic. Plant J.* doi: 10.1016/j.hpj.2024.01.004

Abdul Aziz, M., Sabeem, M., Mullath, S. K., Brini, F., and Masmoudi, K. (2021). Plant group II LEA proteins: Intrinsically disordered structure for multiple functions in response to environmental stresses. *Biomolecules* 11, 1662. doi: 10.3390/biom11111662

Achatz, B., von Ruden, S., Andrade, D., Neumann, E., Pons-Kuhnemann, J., Kogel, K. H., et al. (2010). Root colonization by *Piriformospora indica* enhances grain yield in barley under diverse nutrient regimes by accelerating plant development. *Plant Soil* 333, 59–70. doi: 10.1007/s11104-010-0319-0

Al-Dous, E. K., George, B., Al-Mahmoud, M. E., Al-Jaber, M. Y., Wang, H., Salameh, Y. M., et al. (2011). *De novo* genome sequencing and comparative genomics of date palm (*Phoenix dactylifera*). *Nat. Biotechnol.* 29, 521–527. doi: 10.1038/nbt.1860

Ali, R., Gul, H., Rauf, M., Arif, M., Hamayun, M., Husna Khilji, S. A., et al. (2022). Growth-promoting endophytic fungus (Stemphylium lycopersici) ameliorates salt stress tolerance in maize by balancing ionic and metabolic status. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.890565

Al-Mssallem, I. S., Hu, S., Zhang, X., Lin, Q., Liu, W., Tan, J., et al. (2013). Genome sequence of the date palm *Phoenix dactylifera* L. *Nat. Commun.* 4, 2274. doi: 10.1038/ncomms3274

Athar, H., Zulfiqar, F., Moosa, A., Ashraf, M., Zafar, Z. U., Zhang, L., et al. (2022). Salt stress proteins in plants: An overview. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.999058

Atia, M. A., Abdeldaym, E. A., Abdelsattar, M., Ibrahim, D. S., Saleh, I., Abd Elwahab, M., et al. (2020). *Piriformospora indica* promotes cucumber tolerance against Root-knot nematode by modulating photosynthesis and innate responsive genes. *Saudi J. Biol. Sci.* 27, 279–287. doi: 10.1016/j.sjbs.2019.09.007

Baltruschat, H., Fodor, J., Harrach, B. D., Niemczyk, E., Barna, B., Gullner, G., et al. (2008). Salt tolerance of barley induced by the root endophyte *Piriformospora indica* is associated with a strong increase in antioxidants. *New Phytol.* 180, 501–510. doi: 10.1111/j.1469-8137.2008.02583.x

Bourgis, F., Kilaru, A., Cao, X., Ngando-Ebongue, G. F., Drira, N., Ohlrogge, J. B., et al. (2011). Comparative transcriptome and metabolite analysis of oil palm and date palm mesocarp that differ dramatically in carbon partitioning. *Proc. Natl. Acad. Sci. U.S.A.* 108, 12527–12532. doi: 10.1073/pnas.1106502108

Bouzroud, S., Barbosa, M. A. M., Gasparini, K., Fahr, M., Bendaou, N., Bouzayen, M., et al. (2019). Loss of AUXIN RESPONSE FACTOR 4 function alters plant growth, stomatal functions and improves tomato tolerance to salinity and water deficit. *bioRxiv*, 756387. doi: 10.1101/756387

Cheng, C., Wang, J., Hou, W., Malik, K., Zhao, C., Niu, X., et al. (2021). Elucidating the molecular mechanisms by which seed-borne endophytic fungi, *Epichloë gansuensis*, increases the tolerance of *Achnatherum inebrians* to NaCl stress. *Int. J. Mol. Sci.* 22, 13191. doi: 10.3390/ijms222413191

Cosme, M., Lu, J., Erb, M., Stout, M. J., Franken, P., and Wurst, S. (2016). A fungal endophyte helps plants to tolerate root herbivory through changes in gibberellin and jasmonate signaling. *New Phytol.* 211, 1065–1076. doi: 10.1111/nph.13957

Dinkins, R. D., Nagabhyru, P., Graham, M. A., Boykin, D., and Schardl, C. L. (2017). Transcriptome response of *Lolium arundinaceum* to its fungal endophyte *Epichloë coenophiala*. *New Phytol.* 213, 324–337. doi: 10.1111/nph.14103

Fang, Y., Wu, H., Zhang, T., Yang, M., Yin, Y., Pan, L., et al. (2012). A complete sequence and transcriptomic analyses of date palm (*Phoenix dactylifera* L.) mitochondrial genome. *PloS One* 7, e37164. doi: 10.1371/journal.pone.0037164

Gazara, R. K., Khan, S., Iqrar, S., Ashrafi, K., and Abdin, M. Z. (2020). Comparative transcriptome profiling of rice colonized with beneficial endophyte, *Piriformospora indica*, under high salinity environment. *Mol. Biol. Rep.* 47, 7655–7673. doi: 10.1007/s11033-020-05839-z

Ghaffari, M. R., Ghabooli, M., Khatabi, B., Hajirezaei, M. R., Schweizer, P., and Salekdeh, G. H. (2016). Metabolic and transcriptional response of central metabolism affected by root endophytic fungus *Piriformospora indica* under salinity in barley. *Plant Mol. Biol.* 90, 699–717. doi: 10.1007/s11103–016-0461-z

Gill, S. S., Gill, R., Trivedi, D. K., Anjum, N. A., Sharma, K. K., Ansari, M. W., et al. (2016). *Piriformospora indica*: Potential and significance in plant stress tolerance. *Front. Microbiol.* 7. doi: 10.3389/fmicb.2016.00332

Güimil, S., Chang, H. S., Zhu, T., Sesma, A., Osbourn, A., Roux, C., et al. (2005). Comparative transcriptomics of rice reveals an ancient pattern of response to microbial colonization. *Proc. Natl. Acad. Sci. U.S.A.* 102, 8066–8070. doi: 10.1073/pnas.0502999102

Jisha, S., Sabu, K., and Manjula, S. Multifunctional aspects of *Piriformospora indica* in plant endosymbiosis. (2019). *Mycology* 10, 182–190. doi: 10.1080/21501203.2019.1600063

Jogawat, A., Saha, S., Bakshi, M., Dayaman, V., Kumar, M., Dua, M., et al. (2013). *Piriformospora indica* rescues growth diminution of rice seedlings during high salt stress. *Plant Signal Behav.* 8, e26891. doi: 10.4161/psb.26891

Khatabi, B., Molitor, A., Lindermayr, C., Pfiffi, S., Durner, J., von Wettstein, D., et al. (2012). Ethylene supports colonization of plant roots by the mutualistic fungus *Piriformospora indica*. *PloS One* 7, e35502. doi: 10.1371/journal.pone.0035502

Korkina, L. G. (2007). Phenylpropanoids as naturally occurring antioxidants: From plant defense to human health. *Cell Mol. Biol.* 53, 15–25. doi: 10.1170/T772

Liao, Y., Smyth, G. K., and Shi, W. (2014). featureCounts: an efficient general-purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923–930. doi: 10.1093/bioinformatics/btt656

López-Ráez, J. A. (2016). How drought and salinity affect arbuscular mycorrhizal symbiosis and strigolactone biosynthesis? *Planta* 243, 1375–1385. doi: 10.1007/s00425–015-2435–9

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550. doi: 10.1186/s13059-014-0550-8

Ma, L., Liu, X., Lv, W., and Yang, Y. (2022). Molecular mechanisms of plant responses to salt stress. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.934877

Masmoudi, K., Aziz, M. A., Shamim, A., Sabeem, M., Hazzouri, K. M., and Amiri, K. M. A. (2021). "Metagenomics of beneficial microbes in abiotic stress tolerance of date palm," in *The date palm genome, vol. 2. Compendium of plant genomes*. Eds. J. M. Al-Khayri, S. M. Jain and D. V. Johnson (Springer, Cham), 203–214.

Molitor, A., Zajic, D., Voll, L. M., Pons-K Hnemann, J., Samans, B., Kogel, K. H., et al. (2011). Barley leaf transcriptome and metabolite analysis reveals new aspects of compatibility and *Piriformospora indica*-mediated systemic induced resistance to powdery mildew. *Mol. Plant Microbe Interact.* 24, 1427–1439. doi: 10.1094/MPMI-06-11-0177

Musielak, T. J., Schenkel, L., Kolb, M., Henschen, A., and Bayer, M. (2015). A simple and versatile cell wall staining protocol to study plant reproduction. *Plant Reprod.* 28, 161–169. doi: 10.1007/s00497-015-0267-1

Naser, V., and Shani, E. (2016). Auxin response under osmotic stress. *Plant Mol. Biol.* 91, 661–672. doi: 10.1007/s11103-016-0476-5

Panda, S., Busatto, N., Hussain, K., and Kamble, A. (2019). *Piriformospora indica*-primed transcriptional reprogramming induces defense response against early blight in tomato. *Sci. Hortic.* 255, 209–219. doi: 10.1016/j.scienta.2019.05.031

Qiang, X., Zechmann, B., Reitz, M. U., Kogel, K. H., and Schäfer, P. (2012). The mutualistic fungus *piriformospora indica* colonizes arabidopsis roots by inducing an endoplasmic reticulum stress–triggered caspase-dependent cell death. *Plant Cell* 24, 794–809. doi: 10.1105/tpc.111.093260

R Core Team. (2022). *R: A language and environment for statistical computing* (Vienna: R Foundation for Statistical Computing). Available online at: https://www.R-project.org/ (Accessed February 20, 2024).

Rho, H., Hsieh, M., Kandel, S. L., Cantillo, J., Doty, S. L., and Kim, S. H. (2018). Do endophytes promote growth of host plants under stress? A meta-analysis on plant stress mitigation by endophytes. *Microb. Ecol.* 75, 407–418. doi: 10.1007/s00248-017-1054-3

Sabeem, M., Abdul Aziz, M., Mullath, S. K., Brini, F., Rouached, H., and Masmoudi, K. (2022). Enhancing growth and salinity stress tolerance of date palm using *Piriformospora indica*. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.1037273

Schäfer, P., Pfiffi, S., Voll, L. M., Zajic, D., Chandler, P. M., Waller, F., et al. (2009). Manipulation of plant innate immunity and gibberellin as factor of compatibility in the mutualistic association of barley roots with *Piriformospora indica*. *Plant J.* 59, 461–474. doi: 10.1111/j.1365-313X.2009.03887.x

Shomali, A., Das, S., Arif, N., Sarraf, M., Zahra, N., Yadav, V., et al. (2022). Diverse physiological roles of flavonoids in plant environmental stress responses and tolerance. *Plants* 11, 3158. doi: 10.3390/plants11223158

Shrivasata, P., and Kumar, R. (2015). Soil salinity: a serious environmental issue and plant growth promoting bacteria as one of the tools for its alleviation. *Saudi J. Biologi. Sci.* 22, 123–131. doi: 10.1016/j.sjbs.2014.12.001

Siddiqui, Z. S., Wei, X., Umar, M., Abideen, Z., Zulfiqar, F., Chen, J., et al. (2022). Scrutinizing the application of saline endophyte to enhance salt tolerance in rice and maize plants. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.770084

Trivedi, D. K., Bhatt, H., Johri, A., Tuteja, N., and Bhavesh, N. S. (2013). Sequence-specific ^{1}H, ^{13}C and ^{15}N NMR assignments of Cyclophilin A like protein from *Piriformospora indica* involved in salt stress tolerance. *Biomol. NMR Assign.* 7, 175–178. doi: 10.1007/s12104-012-9404-z

Waller, F., Achatz, B., Baltruschat, H., Fodor, J., Becker, K., Fischer, M., et al. (2005). The endophytic fungus *Piriformospora indica* reprograms barley to salt-stress tolerance, disease resistance, and higher yield. *Proc. Natl. Acad. Sci. U.S.A.* 102, 13386–13391. doi: 10.1073/pnas.0504423102

Xiao, T. T., Raygoza, A. A., Pérez, J. C., Kirschner, G., Deng, Y., Atkinson, B., et al. (2019). Emergent protective organogenesis in date palms: A morpho-devo-dynamic adaptive strategy during early development. *Plant Cell.* 31, 1751–1766. doi: 10.1105/tpc.19.00008

Yang, M., Zhang, X., Liu, G., Yin, Y., Chen, K., Yun, Q., et al. (2010). The complete chloroplast genome sequence of date palm (*Phoenix dactylifera* L.). *PloS One* 5, e12762. doi: 10.1371/journal.pone.0012762

Yin, Y., Zhang, X., Fang, Y., Pan, L., Sun, G., Xin, C., et al. (2012). High-throughput sequencing-based gene profiling on multi-staged fruit development of date palm (*Phoenix dactylifera*, L.). *Plant Mol. Biol.* 78, 617–626. doi: 10.1007/s11103-012-9890-5

Zarea, M. J., Hajinia, S., Karimi, N., Goltapeh, E. M., Rejali, F., and Varma, A. (2012). Effect of *Piriformospora indica* and Azospirillum strains from saline or non-saline soil on mitigation of the effects of NaCl. *Soil Biol. Biochem.* 45, 139–146. doi: 10.1016/j.soilbio.2011.11.006

Zhang, W., Wang, J., Xu, L., Wang, A., Huang, L., Du, H., et al. (2018). Drought stress responses in maize are diminished by *Piriformospora indica*. *Plant Signal Behav.* 13, e1414121. doi: 10.1080/15592324.2017.1414121

Zhou, Y., and Underhill, S. J. (2016). Breadfruit (*Artocarpus altilis*) gibberellin 2-oxidase genes in stem elongation and abiotic stress response. *Plant Physiol. Biochem.* 98, 81–88. doi: 10.1016/j.plaphy.2015.11.012

Zhou, L., Yarra, R., Yang, Y., Liu, Y., Yang, M., and Cao, H. (2022). The oil palm R2R3-MYB subfamily genes *EgMYB111* and *EgMYB157* improve multiple abiotic stress tolerance in transgenic Arabidopsis plants. *Plant Cell Rep.* 41, 377–393. doi: 10.1007/s00299-021-02814-1

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 20 August 2024
DOI 10.3389/fpls.2024.1419227

Check for updates

OPEN ACCESS

EDITED BY
Somashekhar M. Punnuri,
Fort Valley State University, United States

REVIEWED BY
Yiyong Zhao,
Harvard University, United States
Vikas Venu Kumaran,
Indian Agricultural Research Institute (ICAR), India
Zihao Zheng,
Syngenta, United States

*CORRESPONDENCE
Manish K. Vishwakarma
m.vishwakarma@cgiar.org
Arun K. Joshi
a.k.joshi@cgiar.org

RECEIVED 17 April 2024
ACCEPTED 12 July 2024
PUBLISHED 20 August 2024

CITATION
Vishwakarma MK, Bhati PK, Kumar U, Singh RP, Kumar S, Govindan V, Mavi GS, Thiyagarajan K, Dhar N and Joshi AK (2024) Genetic dissection of value-added quality traits and agronomic parameters through genome-wide association mapping in bread wheat (*T. aestivum* L.). *Front. Plant Sci.* 15:1419227. doi: 10.3389/fpls.2024.1419227

COPYRIGHT
© 2024 Vishwakarma, Bhati, Kumar, Singh, Kumar, Govindan, Mavi, Thiyagarajan, Dhar and Joshi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Genetic dissection of value-added quality traits and agronomic parameters through genome-wide association mapping in bread wheat (*T. aestivum* L.)

Manish K. Vishwakarma[1]*, Pradeep K. Bhati[1], Uttam Kumar[2], Ravi P. Singh[3], Sundeep Kumar[4], Velu Govindan[3], Gurvinder Singh Mavi[5], Karthikeyan Thiyagarajan[1], Narain Dhar[1] and Arun K. Joshi[1,3]*

[1]Borlaug Institute for South Asia (BISA), New Delhi, India, [2]Astralyan Agro (OPC) Pvt. Ltd, Shamli, Uttar Pradesh, India, [3]International Maize and Wheat Improvement Center (CIMMYT), Texcoco, Mexico, [4]Indian Council of Agricultural Research (ICAR)-National Bureau of Plant Genetic Resources, New Delhi, India, [5]Department of Plant breeding and genetics, Punjab Agricultural University, Ludhiana, Punjab, India

Bread wheat (*T. aestivum)* is one of the world's most widely consumed cereals. Since micronutrient deficiencies are becoming more common among people who primarily depend upon cereal-based diets, a need for better-quality wheat varieties has been felt. An association panel of 154 *T. aestivum* lines was evaluated for the following quality traits: grain appearance (GA) score, grain hardness (GH), phenol reaction (PR) score, protein percent, sodium dodecyl sulfate (SDS) sedimentation value, and test weight (TWt). In addition, the panel was also phenotyped for grain yield and related traits such as days to heading, days to maturity, plant height, and thousand kernel weight for the year 2017–18 at the Borlaug Institute for South Asia (BISA) Ludhiana and Jabalpur sites. We performed a genome-wide association analysis on this panel using 18,351 genotyping-by-sequencing (GBS) markers to find marker-trait associations for quality and grain yield-related traits. We detected 55 single nucleotide polymorphism (SNP) marker trait associations (MTAs) for quality-related traits on chromosomes 7B (10), 1A (9), 2A (8), 3B (6), 2B (5), 7A (4), and 1B (3), with 3A, 4A, and 6D, having two and the rest, 4B, 5A, 5B, and 1D, having one each. Additionally, 20 SNP MTAs were detected for yield-related traits based on a field experiment conducted in Ludhiana on 7D (4) and 4D (3) chromosomes, while 44 SNP MTAs were reported for Jabalpur on chromosomes 2D (6), 7A (5), 2A (4), and 4A (4). Utilizing these loci in marker-assisted selection will benefit from further validation studies for these loci to improve hexaploid wheat for better yield and grain quality.

KEYWORDS

wheat, GWAS, quality traits, grain protein, MTAS, SNPs, end-use quality

Introduction

Wheat (*Triticum* spp.) is a major staple crop in many countries, including India, and accounts for nearly 30% of global cereal consumption (van Ittersum et al., 2016). Although this main food crop is consumed chiefly as unleavened flatbread (chapati), 15% of the produced yield is used in baking for other bakery items such as bread and cookies. Wheat's value-added characteristics are critical for home consumption and the baking industry (Cappelli and Cini, 2021). Wheat quality is currently described using a variety of metrics, and a single quality parameter can effectively distinguish wheat genotypes of variable qualities (Guzmán et al., 2019). Grain appearance, test weight, grain protein, grain hardness, sedimentation, gluten content, gluten index, iron, zinc, phenol score, and flour extraction are used to categorize wheat varieties suitable for specific end-products such as bread, biscuits, and chapatis. Wheat varieties have been divided into distinct product-specific genotypes based on these qualities in different nations.

The GlutoPeak test is used to forecast the baking qualities of wheat flour, and the association of GlutoPeak indices with several conventional quality measures such as grain hardness (GH), sodium dodecyl sulfate sedimentation value (SDSS value), farinograph, and alveograph has been investigated (Mecitoğlu Güçbilmez et al., 2019). The SDSS test is a quick test to forecast baking quality and gluten strength in wheat (Carter et al., 1999). Low alveograph stability, strength, P/L ratio, protein content, and high alveograph extensibility and biscuit diameter relate to soft endosperm genes in wheat, which are responsible for enhanced biscuit-making capacity. Ma and Baik (2018) reported that soft wheat varieties with low protein content (7.9–9.7%), low sedimentation volume (20.0–32.0 mL), and low damaged starch contents (1.9–3.4%) are desirable for good biscuit-making quality. Various physio-chemical parameters such as grain appearance (GA) score, grain hardness (GH), test weight (TWt), thousand kernel weight (TKW), protein, gluten content and index, SDSS value, phenol test, carotenoids, and diastatic activity are known to have a role in chapati-making quality (Kumar S. et al., 2018). In addition to this, for making a good loaf of bread the combination of elastic gluten with grain protein content of 13% is a prerequisite (Shuey, 1960). Wheat cultivars that have sedimentation values between 35 and 50 cc are typically used to make chapatis, while higher values than that are utilized to make bread (Mecitoğlu Güçbilmez et al., 2019). While GH and diastase activity play a clear role, it was found that phenol score may not be a good indicator of chapatti quality.

Although quality traits are important, bread wheat's grain yield (GY) potential and stress tolerance must be increased to ensure global food security and fulfill future demands. Amid mounting breeding efforts, the low annual rate of GY increase (0.9%) (Ray et al., 2013), the growing menaces of heat and drought stresses on wheat yields (Zampieri et al., 2017), patterns of GY stagnation (Ray et al., 2012), invite the complementation of traditional breeding approaches with genomic tools that can hasten the development of high-yielding and stress-resilient wheat varieties. Wheat GY, however, has remained a challenging trait for genomic breeding due to its quantitative genetic regulation, including numerous loci with minor effects, a shortage of knowledge about the genetic basis of GY, unstable GY quantitative trait loci (QTL) reported in a different environment, epistatic effects, low heritability of GY across environments, and genotype × environment interactions (Jiang et al., 2017). Therefore, to effectively use genetic resources in breeding programs to increase wheat grain production, we must improve our knowledge of the genetic architecture of grain yield and other related attributes.

The molecular basis of complex traits is frequently studied via QTL mapping based on linkage analysis. However, mapping populations such as recombination-inbred lines (RILs) take a long time and much money to create. Furthermore, linkage mapping is based on recent recombination events, resulting in low mapping resolution, and only two alleles from the parents are considered. A genome-wide association study (GWAS) based on linkage disequilibrium (LD) represents an alternate strategy for examining connections between genotype and phenotype with the introduction of high-throughput sequencing technology (Gupta et al., 2020).

A GWAS has various advantages compared with linkage mapping, including a greater resolution and the ability to detect more variation without requiring mapping populations. GWAS has been successfully performed to explore various traits in a range of crops. In wheat, GWAS has been used to investigate grain yield, agronomic traits (Liu et al., 2017), and disease resistance (Liu et al., 2017; Riaz et al., 2018. However, only some studies have focused on quality-related traits in wheat under environmental stresses and grain yield-related traits. Hence, the main goal of this study was to use the mixed-linear model for GWAS of value-added traits and grain yield-related traits using 154 advanced breeding lines of genomic selection nurseries grown at the Ludhiana and Jabalpur Borlaug Institute for South Asia (BISA) sites. We also analyzed the phenotypic distributions of the traits and the statistical correlations between these traits. In addition, we used the KnetMiner to explore the homologous genes in other species with the reported marker trait associations (MTAs) in this study. Based on the available literature, this is the first GWAS using genotyping-by-sequencing (GBS) to examine the stability of value-added quality traits in spring wheat. Our findings provide an understanding of the genetic pathways underlying quality-related attributes.

Materials and methods

Plant material and phenotyping of grain yield and related traits

The panel of 154 selected advanced breeding lines of wheat (Supplementary Table 1) was evaluated in field trials at the BISA research farms, Jabalpur (JBP) (23°14′00.6N and 80°04′40.7E) and Ludhiana (LDH) (30°59′28.74N and 75°44′10.87E). The alpha-lattice experimental design was followed in two replications. The plot size was 5.016 m2, and the lines were sown in six rows, 22 cm apart and 3.8 m in length. The field trials were managed by standard agronomic practices recommended for the locations. Fertilizer was applied with the proportions of 150 N/60 P/40 K kg/ha at Ludhiana

and 120 N/60 P/40 K kg/ha at Jabalpur as per the wheat growing zone recommendations.

During the 2017–18 crop season, the lines were phenotyped and evaluated across the location for five traits such as days to heading (DTHD) and days to maturity (DAYSMT). DTHD and DAYSMT were measured as the total number of days from sowing to when 75% of plants had either spike emergence or matured, respectively. Plant height (PH) was recorded from the plant's base to the tip of the spike (excluding awns). Thousand kernel weight (TKW) and grain yield (GY) were measured per plot.

Estimation of value-added quality traits

Quality traits data was recorded at the Wheat Quality Laboratory, Punjab Agricultural University (PAU), Ludhiana, Punjab for six grain quality parameters including protein percent, TWt, GA, PR score,SDSS value, and GH.

Using an Infratec 1226 Cold Grain Analyzer and the AACCI standard procedure, protein percentage was measured non-destructively at 12% moisture basis. The instrument uses near-infrared light transmitted through the grains. The results are displayed as % protein content as per calibration.

TWt, also known as hectoliter mass, measures the volume of grain per unit. Hectoliter weight was determined using a Tecator model FP Auto 680 by taking wheat grains in a 100 mL measuring cylinder; the sample was weighed, and the hectoliter weight was expressed as kg ha−1 (AACC, 2000).

Subsequently, we measured the GA score through direct visualization based on the grain's size, shape, color, and luster. It was evaluated subjectively out of a maximum score of 10. The phenol reaction score was evaluated by soaking about 100 grains overnight in 1% phenol solution. The grains were assessed for the extent of darkness out of a score of 10, half an hour after draining off the phenol solution.

The SDSS test was used since it is a simple, small-scale method that estimates wheat gluten strength quickly. The SDSS test was carried out according to Nakamura et al., (2012). The SDS-lactic acid solution was prepared by dissolving 20 g of SDS in 1 L of distilled water and adding 20 mL of stock diluted lactic acid solution (one-part lactic acid plus eight parts distilled water volume by volume). Six grams of the whole meal sample were placed in a stoppered, graduated cylinder with 50 mL of water. The samples were mixed, hydrated for 2 min, remixed, and then hydrated for another 2 min. SDS–lactic acid solution (50 mL) was added to each sample, and the contents were mixed by inverting the tubes four times. The contents were allowed to settle, and the sedimentation height (mL) was recorded. If the value was more than 60 mL, it was considered as strong gluten wheat; from 30 to 60, it was medium strong, and if less than 30 mL, it was weak.

The GH was measured using the grain hardness tester supplied by M/S Ogawa Seiki Co. Ltd., Japan, by crushing randomly taken ten grains one by one, considering the weight, diameter, and moisture of the grain. The mean force (kg) required to crush the grain was recorded (Heo and Sherman, 2013).

Statistical analysis

The experimental design in each environment was an alpha-lattice with two replications per environment/location. The best linear unbiased prediction (BLUP) values were obtained through META-R v6.03 (Alvarado et al., 2020). All effects are considered random for calculating the BLUP and broad-sense heritability. The correlation matrix between the BLUP values of studied traits was computed and visualized with the 'corrplot' package in the R software.

Genotyping, linkage disequilibrium

Genomic DNA of the lines was isolated from 15 days-old seedling leaves using a standard cetyltrimethyl ammonium bromide (CTAB) method (Doyle and Doyle, 1987). DNA concentration was quantified using the Quant-iT PicoGreen dsDNA assay (Life Technologies Inc., NY) and normalized to 20 ng/µl. The panel of 154 lines was genotyped using the GBS method (Poland et al., 2012). The single nucleotide polymorphisms (SNPs) were called using the TASSEL (Trait Analysis by association Evolution and Linkage) version V5.3.1 GBS pipeline (Glaubitz et al., 2014). Marker polymorphisms were found using a minor allele frequency of 0.01, which resulted 13,082,477 GBS tags. Among these, 68.98% were aligned to RefSeq v1.0 using Bowtie2 (Langmead and Salzberg, 2012) with assembly of Chinese Spring (IWGSC, 2018). After filtering the tags as described by Juliana et al. (2019), we found 89,863 SNPs. Then, these markers were filtered in panel, and those with more than 60% missing data, a minor allele frequency of less than 5%, or heterozygosity of less than 10% were eliminated. Similarly, the markers and lines with a total missing data percentage of more than 50% were eliminated and 18,351 polymorphic markers were used for all the subsequent analyses. LD analysis was performed using TASSEL V5.3.1 software (Bradbury et al., 2007) using the markers with known positions from the 18,351 polymorphic markers. The LD was estimated as squared allele frequency correlations (R^2). P-values <0.01 for each pair of loci and Bonferroni correction <0.2 were considered significant.

Genome-wide association scans for grain quality and agronomic traits

Six grain quality parameters (GA Score, GH, PR Score, Protein %, SDSS Value, and TWt), and five agronomic traits (DTHD, DAYSMT, PH, GRYLD, and TKW) from both the locations (Ludhiana and Jabalpur) were considered for a GWAS using 18351 polymorphic GBS markers. GWAS analysis was performed with TASSEL V5.3.1 software (Bradbury et al., 2007) using a Mixed Linear Model (MLM). Population structure was used as a fixed effect in the model's fitting, while kinship was used as a random effect that was considered by the first two principal components (Patterson et al., 2006; Price et al., 2006).

10.3389/fpls.2024.1419227

Detection of marker trait associations for quality traits

Associations of GA Score, GH, PR Score, Protein %, SDSS Value, and TWt, with candidate loci were identified. We obtained the p-values to determine the significance of the association of traits with the markers and the percent variance explained (PVE), which predicted the extent of the QTL effects. The Manhattan plots for grain quality traits were generated in the GWAS, indicating the most significant associations with a –log10 (P value) greater than 3, along with the Bonferroni correction threshold (we used the Bonferroni correction for multiple testing with an α level of 0.01 for the quality traits and a relaxed α level of 0.20 for all the other datasets) and quantile-quantile (Q-Q) plots.

Prediction of candidate gene and modeling of homology

The ENSEMBL Wheat database and the International Wheat Genome Sequencing Consortium (IWGSC) RefSeq v1.1 annotations were used to find candidate genes related to the stable loci discovered in this investigation. To find candidate genes, regions within the 1 Mbp window of the localized stable MTA were also chosen. For the gene network analysis and homology finding, an open-source online software, Knetminer, was used at: https://knetminer.org (accessed on Oct 28, 2022) (Hassani-Pak et al., 2021).

Results

Phenotypic variation and heritability grain quality traits and agronomic traits

A range of variation for all grain quality traits was reported in the advanced breeding lines of the spring wheat panel. The Protein % ranged from 8.50 to 12.35, with a mean of 10.35 and a CV of 8.61%. Similarly, TWt, GA Score, PR Score, SDSS values, and GH ranged from 66.50 to 78.00, 4.0 to 6.0, 2.20 to 5.50, 29.0 to 50.0, and 7.70 to 12.0 respectively (Table 1).

Concerning the quantitative traits analysis, a range of variation was observed for GRYLD and other yield related traits at both the locations (Ludhiana and Jabalpur). The broad-sense heritability for the traits under consideration ranged from 0.40 to 0.91 (Table 2). The highest broad-sense heritability (0.91) was observed for DTHD at Jabalpur, while the same for GRYLD at Ludhiana and Jabalpur were recorded as 0.69 and 0.60, respectively. Similarly, nearly stable and high heritability were observed for TKW at Ludhiana (0.82) and Jabalpur (0.74). An excellent yielding line (GID: 6692345; SOKOLL/3/PASTOR//HXL7573/2*BAU/4/SOKOLL/WBLL1) with a yield of >7.0 t/h was observed for the Ludhiana location, while there were three lines (GID: 6681676, QUAIU#1/SUP152; GID: 6681793, ND643/2*WBLL1/4/WHEAR/KUKUNA/3/C80.1/3*BATAVIA//2*WBLL1 and GID: 6681817 SUP152/QUAIU#2) were >7.8 t/h for the Jabalpur location. In grain quality traits, the lines with GID 6568703 (PRL/2*PASTOR/4/CHOIX/STAR/3/HE1/3*CNO79//2*SERI*2/5/CHONTE), 6692267 (PASTOR//HXL7573/2*BAU/3/ATTILA/3*BCN/4/SOKOLL/3/PASTOR//HXL7573/2*BAU), and 6692345 (SOKOLL/3/PASTOR//HXL7573/2*BAU/4/SOKOLL/WBLL1) had 12.35, 12.32, and 12.28 protein %, respectively. The lines (GID: 6684333, SWSR22T.B./2*BLOUK #1//WBLL1*2/KURUKU) had high test weight value (78) while three lines (GID: 6568578, KIRITATI/4/2*SERI.1B*2/3/KAUZ*2/BOW//KAUZ/5/2*SUP152, GID: 6568703, PRL/2*PASTOR/4/CHOIX/STAR/3/HE1/3*CNO79//2*SERI*2/5/CHONTE and GID: 6684107, MUTUS*2/HARIL #1) had high grain hardness of 11.8, 11.8, and 12.0, respectively.

TABLE 1 Descriptive statistics of various quality traits of wheat.

Traits	Range		Mean	SD	CV%
	Minimum	Maximum			
Protein %	8.50	12.35	10.35	0.89	8.61
TWt	66.50	78.00	73.95	2.00	2.70
GA Score	4.00	6.00	5.41	0.30	5.51
PR Score	2.20	5.50	3.25	0.42	12.82
SDSS Value	29.00	50.00	40.03	4.46	11.15
GH	7.70	12.00	9.60	0.97	10.06

Standard deviation (SD), Coefficient of variance (CV%); Grain Appearance Score (GA Score), Grain Hardness (GH), Phenol Reaction Score (PR Score), Protein Percentage (Protein %), SDS Sedimentation Value (SDSS Value), Test Weight (TWt).

Correlations between agronomic and quality traits of both locations

Location-wise correlation among the agronomic traits was analyzed. For Ludhiana and Jabalpur, the suffixes 'L' and 'J', respectively, have been added to the trait names. GRYLD_L

TABLE 2 Variability analysis of various yield-related agronomic traits at two locations.

Loc	Traits	H^2	G Var	R Var	G Mean	LSD	CV
LDH	DTHD	0.83	6.53	2.63	102.11	3.02	1.59
	DAYSMT	0.68	1.49	4.77	146.80	2.70	1.49
	PH	0.40	2.51	5.13	93.28	2.18	2.37
	TKW	0.82	12.18	5.18	39.41	4.13	5.78
	GRYLD	0.69	0.40	0.37	6.25	1.00	9.68
JBP	DTHD	0.91	9.70	2.01	75.44	2.77	1.88
	DAYSMT	0.62	3.75	4.67	118.39	3.39	1.83
	PH	0.47	5.12	11.62	100.18	4.61	3.40
	TKW	0.74	10.52	7.49	46.40	4.68	5.90
	GRYLD	0.60	0.13	0.17	7.13	0.65	5.86

Loc, location; Env, Environment; H2, heritability; G Var, genotypic variance; R Var, residual variance; LSD, least significant difference; CV, critical variance; LDH, Ludhiana; JBP, Jabalpur. DTHD, days to heading; DAYSMT, days to maturity; GRYLD, grain yield; TGW, thousand-grain weight.

showed a positive correlation with TKW_L, DTHD_L, PH_L, and DAYSMT_L with values of 0.39, 0.17, 0.12, and 0.10, respectively (Figure 1; Supplementary Table 2). TKW_L showed a positive correlation with PH_L with the value of 0.11 and a negative correlation with DAYSMT_L and DTHD_L with -0.18 and -0.09 values, respectively. GRYLD_J showed a positive (0.04) correlation with DAYSMT_J and TKW_J, while it had a negative correlation with DTHD_J, and PH_J with -0.21 and -0.13, respectively. TKW_J showed a positive correlation with PH_J with a value of 0.30 and a negative correlation with DAYSMT_J and DTHD_J with -0.14 and -0.10 values, respectively.

The correlation between the quality traits revealed that there were positive correlations between all the traits. For example, Protein % has a high correlation with the TWt, GA score, PR score, SDSS value, and GH with the values of 0.62, 0.46, 0.39, 0.59, and 0.36, respectively. Similarly, TWt had a high correlation with the GS score, PR score, SDSS value, and GH with the values of 0.92, 0.45, 0.64, and 0.66, respectively. We observed a positive correlation between SDSS value and HG (0.51). A similar pattern was observed for the GA score with the PR score (0.39), SDSS value (0.56), and GH (0.66) (Supplementary Table 1). There was a significant correlation between the GA score and PR score, SDSS value, and GH with values of 0.39, 0.56, and 0.66, respectively. In addition, the PR score correlated well with SDSS and GH with values of 0.33 and 0.34. In addition, SDSS value had a high correlation with GH with a value of 0.51.

The correlation between GRYLD and quality traits across sites (Ludhiana and Jabalpur) elucidated that, GRYLD_L has positive correlation with Protein%, TWt, GA Score, SDSS Value, and GH (0.11, 0.15, 0.23, 0.23 and 0.17, respectively) and showed a negative correlation with the PR Score (-0.18). Furthermore, GRYLD_J had very low correlation with GA Score (0.02) and PR Score (0.12).

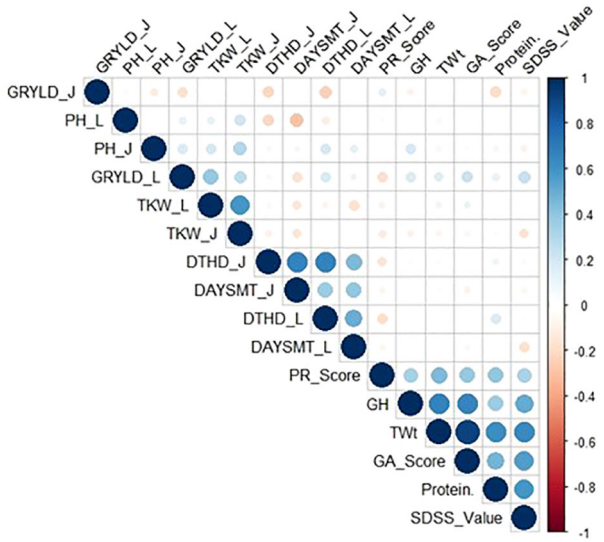


FIGURE 1
Correlation of agronomic traits in wheat lines. Phenotypic correlations between days to heading (DTHD), days to maturity (DAYSMT), plant height (PH), grain yield (GRYLD), thousand kernel weight (TKW), Ludhiana (_L), Jabalpur (_J), grain appearance score (GA Score), grain hardness (GH), phenol reaction score (PR Score), protein percentage (Protein %), SDS sedimentation value (SDSS Value), and test weight (TWt). The upper and lower 95% confidence intervals are included in parenthesis below the correlation value.

Marker densities and population structure

Marker densities of all 18351 GBS markers utilized, aligned to RefSeq v1.0, showed that the telomeric and sub-telomeric regions had higher densities than the centromeric regions in all chromosomes (Figure 2). The B-genome has the highest number of markers (48.8%), followed by the A-genome (36.3%) and the D-genome (13.6%). The SNPs in high linkage disequilibrium with one another are reflected by the red area and are consequently inherited together (Figure 3). Population structure analysis of all the 154 lines in this study indicated moderate population structure, high diversity, and relatedness between the lines across the locations. The first two principal components plot, PC1 and PC2, explain 6.9% and 5.4% of the variation, respectively (Figure 4).

Detection of marker trait associations for grain quality traits

For the quality traits, a total of 55 significantly (P <0.001 and Bonferroni correction cut-off value of 0.2) associated SNPs were detected (Table 3, Figure 5). For GA Score, two SNPs, S5A_671478896 and S7B_613779914, were identified on chromosomes 5A and 7B that explained 18% and 27% of phenotypic variation, respectively. For GH, 13 MTAs were detected on 1A, 1D, 2B, 3A, 4A, 4B, and 7B with 9 – 13% PVE. Interestingly, seven MTAs were on chromosome 7B. Furthermore, SNPs S7B_689902344 and S2B_13408810 had 13% and 12% PVE, respectively. For PR score, five MTAs were detected only on a single chromosome, 2A, with 20–22% PVE, and two SNPs, S2A_707007872 and S2A_707063443, had 22% PVE.

For the Protein %, ten MTAs were obtained on chromosomes 1A, 1B, 3B, 5B, and 7A with 10- 13% PVE, whereas 3A consisted of five MTAs alone. SNP S7A_13179057 had the highest PVE at 13% and S3B_720255460 had 12% PVE. Furthermore, 13 MTAs were found for the SDSS value on chromosomes 1A, 1B, 2B, 4A, and 7B with a range of 10–14% PVE, with a maximum of six MTAs on the 1A chromosome. SNP S1A_49281757 had the highest PVE, 14%, followed by the SNP S1A_49239494 with 12% PVE on the same chromosome. For the TWt, 12 MTAs were reported on chromosomes 1A, 2A, 2B, 3A, 3B, 6D, 7A, and 7B with 10–14% PVE.SNPs S2A_48176393 and S7A_506298541 had the highest (14%) PVE.

Detection of marker trait associations for agronomic traits

For the Ludhiana location, a total of 20 MTAs were detected for all the agronomic traits (Table 4). Three MTAs, S4A_84900641, S4B_664526264, and S5A_470192586, on chromosomes 4A, 4B, and 5A, respectively, were found for the DAYSMT_L with a range of 10–11% of PVE (Figures 6, 7). Only two MTAs, S4D_456260804a and S4D_457212141, on single chromosome 4D were obtained for

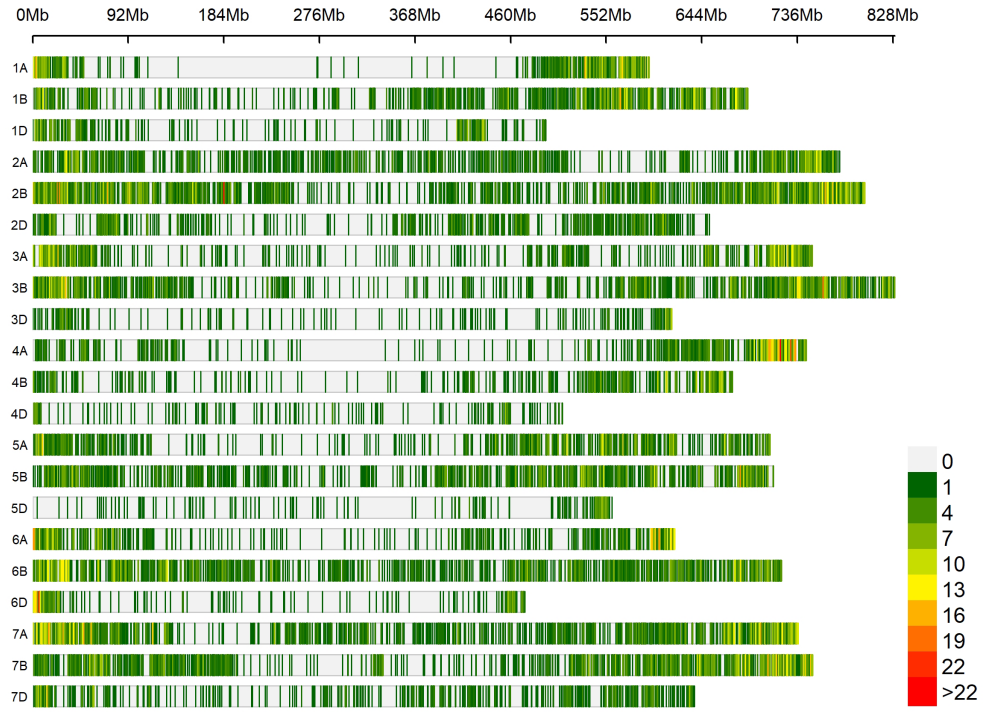


FIGURE 2
Densities of 18,351 genotyping-by-sequencing markers in the reference bread wheat genome (RefSeq v1.0). The color key with marker densities indicates the number of markers within a window size of 1 Mb.

the DTHD_L with 14% PVE. Five MTAs were found on chromosomes 2B, 2D, 7B, and UN (non-confirmed location) for PH_L with a range of 10–14% PVE. SNP SUN_32203753 had the highest PVE of 14%, and SNPs S2B_49523499, S2D_69502623, and S2D_67201447 had 11% PVE. Furthermore, eight MTAs were reported for TKW_L with a range of 10–14% PVE on the 2B, 3B, 5B, 7A, and 7D chromosomes where four MTAs shared the 7D chromosome alone. SNPs S7D_450126108 and S2B_565059870 showed a high PVE of 14% and 13%, respectively.

For the Jabalpur location, 44 MTAs were reported for all the agronomic traits (Table 5; Figures 6, 7). We found three MTAs for DAYSMT_J on chromosome 7A (S7A_652922600), 3A (S3A_1395864), and 3D (S3D_553430673) which had a maximum of 13% and 11% PVE, respectively, while the location of one MTA that explains 11% of phenotypic variation was not known (SUN_37992374). Similarly, two MTAs for DTHD_J were found on chromosomes 1B (S1B_511789038), with 15% PVE, and 4B (S4B_525119244), with 14% PVE, while the location of one MTA that explained up to 14% of phenotypic variation was not known (SUN_37992374). Plant height had ten MTAs with a maximum of four on the 2A chromosome, whereas the highest PVE, 27%, was obtained on chromosome 3D (S3D_18502122), followed by 26% PVE on both 1A (S1A_517587620) and 2A (S2A_458243706). For TKW_J, a total of 16 MTAs was observed; a maximum of six were on chromosome 2D, followed by four on Chromosome 4A, and three on chromosome 7A. The highest PVE (13%) was obtained for the SNP S7A_24940330, followed by S6B_393496773 with 12% PVE; the rest of the MTAs were in the range of 10–11% for PVE.

Detection of marker trait associations for grain yield

For grain yield, only two MTA were detected for the Ludhiana location, SNPs S1A_27512785 (Chromosome 1A) and S4D_75146074 (Chromosome 4D), with 13 and 14% of PVE, respectively (Figure 7). At the Jabalpur location, 11 MTAs were obtained for GRYLD with a range of 10–13% PVE. There was a maximum of three MTAs on the chromosome 5A, followed by two on the chromosome 1A; the rest were on single chromosomes such as S1B_460449239 on 1B, S3A_531631224 on 3A, S3B_584821928 on 3B, S4B_648759658 on 4B, S7A_588008821 on 7A, and S7B_707946706 on 7B. Two SNPs, S1A_10214692 on 1A and S5A_556823005, on 5A, had the highest phenotypic variance with 13% and 12% PVE, respectively.

Candidate gene prediction and associated network

A total of 116 SNPs were physically mapped to IWGSC RefSeqv1.1 with high confidence. To identify the putative candidate genes, the 1Mb flanking region of the mapped SNPs was annotated using EnsemblPlant BioMart. This led to the identification of 19 SNPs overlapped by candidate genes (Table 6). Based on the literature survey and current findings, 19 SNPs were considered as novel, and were associated with the following traits: protein %, SDS, PH, DTHD, PR Score, TWt,

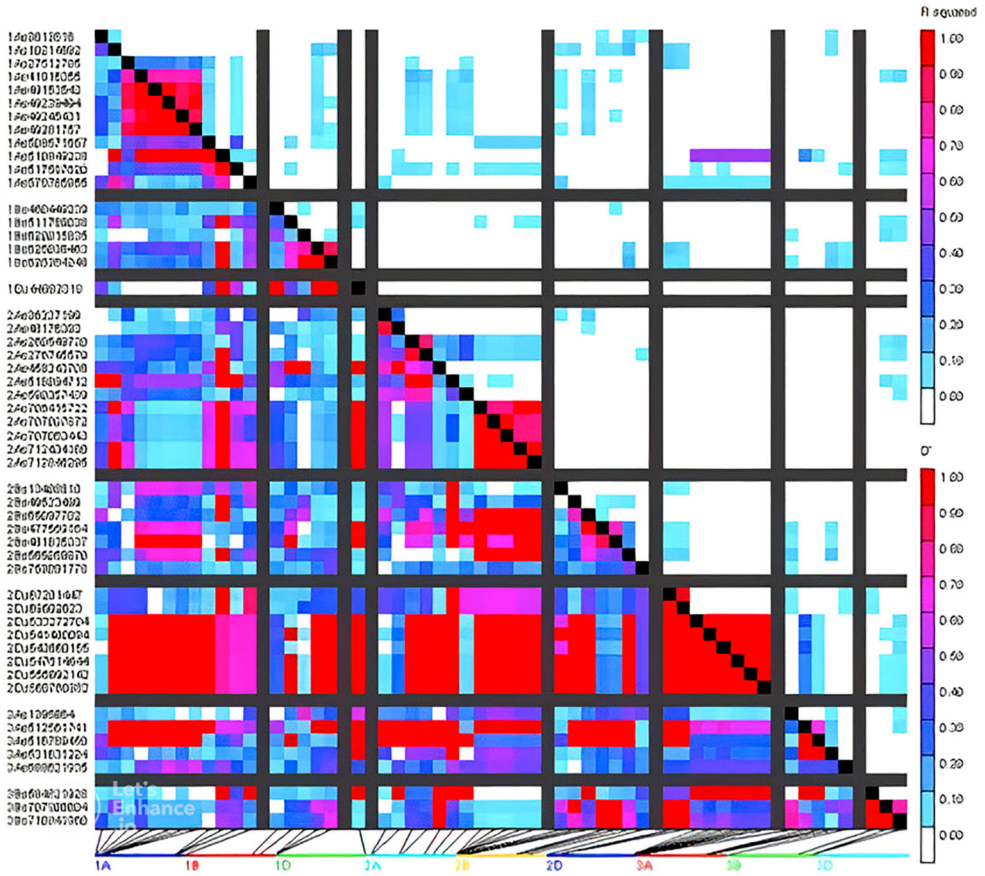


FIGURE 3
Patterns of LD blocks (right) of GWAS results indicating the position of candidate genes and/or QTL regions associated with grain quality traits and agronomic traits.

TKW, GH, DMT, and GYLD. The validation results in KnetMiner network showed that the SNPs for SDSS, such as S1A_49153543, S1A_49281757, S2B_477569164, and S4A_740926925 overlapped with *TraesCS1A02G066900*, *TraesCS1A02G067300*, *TraesCS2B02G333900*, and *TraesCS4A02G491100*.

The SNPs were associated with coding proteins such as SPD1 (involved in plastid development during early seedling growth); RRP5 (role in alternative regulation in plants); DMAS1-B, NRAMP2 and NAAT2-D (Fe/Zn transport and accumulation in grain); and Thioredoxin-like_sf (Redox regulation) (Supplementary Figure 1). The SNPs for TKW (S3A_512561741, S3B_739166411, S4A_679160910, and S7D_476139586) overlapped with *TraesCS3A02G284100*, *TraesCS3B02G494600*, *TraesCS4A02G406300*, and *TraesCS7D02G367800*. These *Traes IDs* code proteins such as GAUT (involved in pectin and xylan biosynthesis); COG_su4 (mediates the proper glycosylation of proteins trafficking through the Golgi apparatus); DHNA_phytyltransferase_MenA (involved in 2-carboxy-1,4-naphthoquinone phytyltransferase); and MIP (These channel proteins function in water, small carbohydrate (e.g., glycerol), urea, NH3, CO2, and possibly ion transport). Likewise, the SNPs for protein %, viz. S3B_720255460, S3B_728890092, and S7A_13179057, were found to be associated with *TraesCS3B02G471800*, *TraesCS3B02G481200*, and *TraesCS7A02G031700* respectively, which code for LRR_dom_sf/NB-ARC (involved in a variety of biological processes); F-box-like_dom_sf/F-box_dom (present in numerous proteins with a bipartite structure); and Aminoacyl-tRNA synthetase (These channel proteins function in water, small carbohydrate (e.g., glycerol), urea, NH3, CO2 and possibly ion transport, by an energy independent mechanism). Similarly, for TWt, PH, GYLD, and DAYSMT, we found 2, 2, 2, and 1 overlapped genes, respectively (see detail in Table 6).

Discussion

The performance of a wheat crop should not be judged only from the angle of grain yield as it has several end-product qualities that determine the market value with different value-added parameters. We must explore the different combinations of additional value-added quality parameters to select for desirable end-use quality. Molecular markers linked to the desirable traits is a holistic approach and can be applied in molecular breeding as a tool to identify varieties and lines at any crop development stage. This study attempted the high-resolution genetic dissection of quality, yield, and agronomic variables in spring wheat to find new valuable alleles in genotypes.

The heritability of GRYLD and TKW was good, while for the phenological traits, it was high for DTHD and moderate for

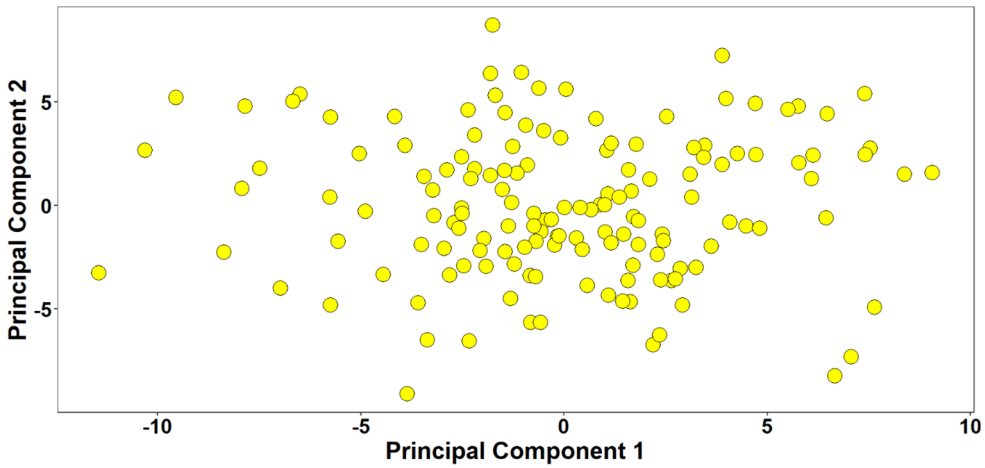


FIGURE 4
Population structure analysis of 154 lines. The plot of the first two principal components explaining 6.9% and 5.4% of the variation, respectively indicated weak population structure with high relatedness between the lines.

TABLE 3 Genome wide significant associations (R2) of single nucleotide polymorphisms (SNPs) with quality traits in wheat.

Trait	SNP	Chr	Alleles	Position (Mb)	p-Value	add_effect	PVE%
GA Score	S5A_671478896	5A	G/A	671.48	6.17E-06	-1.01E-01	18
	S7B_613779914	7B	G/T	613.78	3.06E-08	3.11E-03	27
GH	S1A_3613616	1A	T/C	3.61	2.85E-04	1.61E-01	11
	S1D_44982319	1D	C/T	44.98	7.71E-04	-7.54E-02	10
	S2B_13408810	2B	A/C	13.41	1.87E-04	-2.28E-01	12
	S3A_688621935	3A	A/G	688.62	7.48E-04	1.18E-01	10
	S4A_586697513	4A	C/A	586.7	7.70E-04	5.62E-01	10
	S4B_230165089	4B	C/T	230.17	8.62E-04	-2.45E-01	9
	S7B_689902344	7B	G/A	689.9	7.99E-05	5.19E-01	13
	S7B_702516379	7B	T/C	702.52	2.55E-04	8.81E-01	11
	S7B_687596490	7B	C/T	687.6	2.69E-04	-5.34E-01	11
	S7B_702554771	7B	A/C	702.55	2.86E-04	-8.81E-01	11
	S7B_703152055	7B	C/A	703.15	4.18E-04	6.25E-01	10
	S7B_689673455	7B	A/G	689.67	6.48E-04	-4.74E-01	10
	S7B_689968561	7B	G/C	689.97	9.12E-04	5.17E-01	9
PR Score	S2A_707007872	2A	T/C	707.01	1.27E-07	-	22
	S2A_707063443	2A	T/C	707.06	1.27E-07	-	22
	S2A_706416722	2A	C/T	706.42	3.06E-07	-	20
	S2A_712434160	2A	T/C	712.43	3.06E-07	-	20
	S2A_712846295	2A	G/C	712.85	3.06E-07	-	20
Protein %	S1A_508571657	1A	A/G	508.57	4.27E-04	-2.81E-01	11
	S1B_620015835	1B	G/A	620.02	9.26E-04	-4.73E-01	10

(Continued)

10.3389/fpls.2024.1419227

TABLE 3 Continued

Trait	SNP	Chr	Alleles	Position (Mb)	p-Value	add_effect	PVE%
	S3B_720255460	3B	C/T	720.26	2.17E-04	6.07E-01	12
	S3B_707906604	3B	A/G	707.91	4.22E-04	3.48E-01	11
	S3B_715945072	3B	G/T	715.95	5.56E-04	3.96E-01	11
	S3B_710841960	3B	A/G	710.84	8.53E-04	3.54E-01	10
	S3B_728890092	3B	A/C	728.89	8.60E-04	4.78E-01	10
	S5B_46768581	5B	T/C	46.77	6.35E-04	1.82E-01	11
	S7A_13179057	7A	G/A	13.18	1.65E-04	-4.27E-01	13
	S7A_15198988	7A	C/T	15.2	8.77E-04	-3.57E-02	10
SDSS Value	S1A_49281757	1A	T/C	49.28	5.86E-05	2.87E+00	14
	S1A_49239494	1A	T/A	49.24	2.79E-04	2.80E+00	12
	S1A_49245431	1A	G/A	49.25	3.71E-04	2.70E+00	11
	S1A_41916355	1A	A/G	41.92	4.19E-04	-2.27E+00	11
	S1A_49153543	1A	T/G	49.15	5.04E-04	2.62E+00	11
	S1A_510849238	1A	G/A	510.85	6.65E-04	-2.31E+00	11
	S1B_625036463	1B	A/G	625.04	7.50E-04	-1.73E+00	10
	S1B_625364248	1B	G/T	625.36	9.13E-04	-1.58E+00	10
	S2B_477569164	2B	T/C	477.57	3.96E-04	-1.63E+00	11
	S2B_481835337	2B	G/A	481.84	4.30E-04	-1.96E+00	11
	S2B_65097702	2B	G/A	65.1	6.34E-04	2.35E+00	11
	S4A_740926925	4A	C/A	740.93	9.91E-04	-1.30E+00	10
	S7B_723395908	7B	C/G	723.4	7.31E-04	-8.09E-01	10
TWt	S1A_579785955	1A	A/G	579.79	4.32E-04	8.01E-01	11
	S2A_48176393	2A	C/T	48.18	7.72E-05	3.13E-01	14
	S2A_518094712	2A	C/T	518.09	3.87E-04	1.27E+00	12
	S2A_36227199	2A	G/C	36.23	5.41E-04	-3.66E-01	11
	S2B_753091778	2B	G/C	753.09	9.66E-04	8.16E-01	10
	S3A_516789450	3A	G/C	516.79	1.91E-04	-4.34E-01	13
	S3B_768723701	3B	G/A	768.72	7.02E-04	5.95E-01	11
	S6D_3132722	6D	A/G	3.13	5.44E-04	6.73E-01	11
	S6D_3447720	6D	A/G	3.45	9.43E-04	6.97E-01	10
	S7A_506298541	7A	A/G	506.3	7.51E-05	-1.40E-01	14
	S7A_699093945	7A	A/G	699.09	5.32E-04	-5.44E-01	11
	S7B_613779914	7B	G/T	613.78	2.78E-04	2.38E-01	12

Grain Appearance Score (GA Score), Grain Hardness (GH), Phenol Reaction Score (PR Score), Protein Percentage (Protein %), SDS Sedimentation Value (SDSS Value), Test Weight (TWt), Percent Variance Explained (PVE).

DAYSMT and PH. This implies that the phenotypic measurements were of very high quality and that the attributes had a high degree of predictive power. It was reported (Maphosa et al., 2014; Würschum et al., 2018) that GRYLD, a highly quantitative and environmentally sensitive trait, showed significant variation among environments. We also found that agronomic traits significantly contributed to variance explanation and that their heritability was lower than that of other factors, indicating a considerable G×E impact on GRYLD. Therefore, moderate heritability values for GRYLD were anticipated, given that multiple genes govern it. The lower sowing density with smaller plots may also impact the low heritability and yield variances (Thorwarth et al., 2017; Bhatta et al., 2018). The two

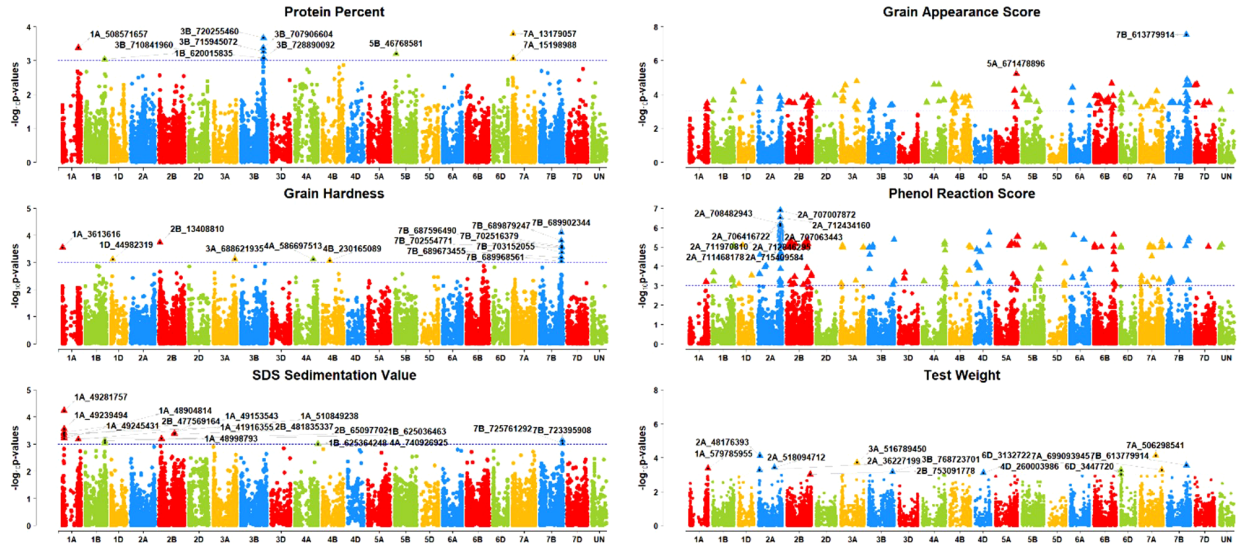


FIGURE 5
Manhattan plots of GWAS results for grain quality traits (grain appearance score (GA Score), grain hardness (GH), phenol reaction score (PR Score), protein percentage (Protein %), SDS sedimentation value (SDSS Value), and test weight (TWt).); Threshold = −log10(p−value) > 3.

TABLE 4 Genome wide significant associations (R2) of single nucleotide polymorphisms (SNPs) with agronomic traits in wheat at Ludhiana.

Trait	SNP	Chr	Allele	Position (Mb)	p-Value	Add effect	PVE%
DAYSMT_L	S4A_84900641	4A	C/T	84.901	8.87E-04	0.76	10
	S4B_664526264	4B	G/A	664.526	6.20E-04	-0.45	11
	S5A_470192586	5A	T/G	470.193	7.63E-04	0.38	10
DTHD_L	S4D_456260804	4D	C/G	456.261	6.23E-05	1.14	14
	S4D_457212141	4D	C/T	457.212	8.54E-05	1.13	14
GRYLD_L	S1A_27512785	1A	G/A	27.513	9.09E-05	-0.23	13
	S4D_75146074	4D	C/T	75.146	7.72E-05	0.55	14
PH_L	S2B_49523499	2B	C/A	49.523	7.07E-04	-0.17	11
	S2D_69502623	2D	T/C	69.503	7.81E-04	-0.14	11
	S2D_67201447	2D	T/G	67.201	7.85E-04	-0.14	11
	S7B_684596722	7B	T/C	684.597	8.40E-04	-0.27	10
	SUN_32203753	UN	A/T	32.204	1.05E-04	0.04	14
TKW_L	S2B_565059870	2B	T/C	565.06	1.70E-04	0.91	13
	S3B_739166411	3B	C/T	739.166	4.87E-04	-0.33	11
	S5B_383209462	5B	A/G	383.209	9.60E-04	1.11	10
	S7A_52015267	7A	C/T	52.015	7.73E-04	-0.25	11
	S7D_450126108	7D	G/A	450.126	1.12E-04	-1.49	14
	S7D_566354436	7D	T/G	566.354	5.79E-04	-1.93	11
	S7D_365824018	7D	A/G	365.824	8.77E-04	1.68	10
	S7D_476139586	7D	C/T	476.14	9.28E-04	2.41	10

DTHD, days to heading; DAYSMT, days to maturity; GRYLD, grain yield; TKW, thousand-kernel weight; PVE, percent variance explained.

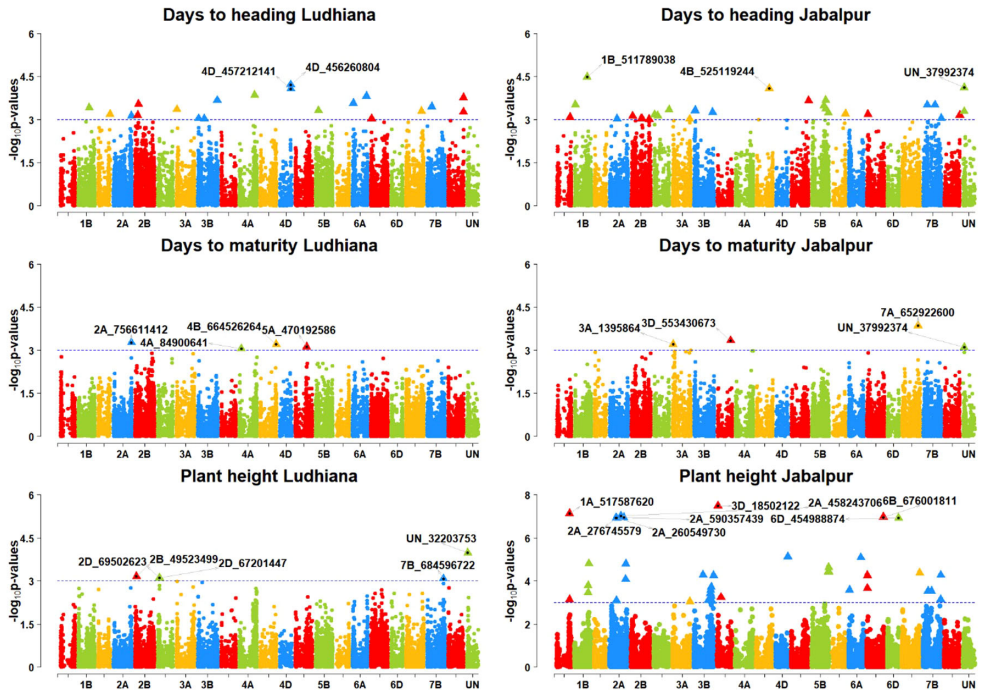


FIGURE 6
Manhattan plots of GWAS results for agronomic traits days to heading (DTHD), days to maturity (DAYSMT), and grain yield (GRYLD); Ludhiana (LDH), Jabalpur (JBP); Threshold = –log10(p–value) > 3.

locations used in this study have very distinct climates. Due to the high ambient temperature, the growing season is comparatively shorter in Jabalpur than in Ludhiana, which has a significantly colder environment with longer growing seasons (Mondal et al., 2016).

Grain quality is a cumulative effect of several traits such as grain protein content, grain hardness, GA score, PR score, SDSS value, and test weight. There is a perception that for good end-products and chapati-making quality, there is a specific combination of the desired grain quality features. To ascertain which combination of quality has what relationship with the others, we estimated correlations between the grain quality of the advanced breeding lines. Likewise, we proceeded with the correlation study to elucidate the correlation between the agronomic and quality traits. There are only a few reports where correlations between the end-use quality traits such as protein %, TWt, GA Score, PR Score, GH, and SDSS value in spring wheat in multi-environment were studied (Guzmán et al., 2016; Ibba et al., 2020; Tsenov et al., 2020).

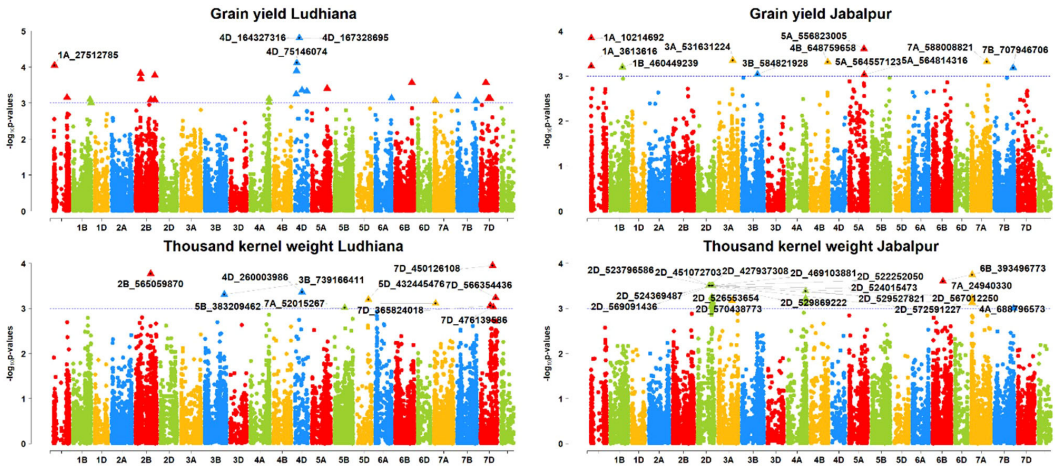


FIGURE 7
Manhattan plots of GWAS results for agronomic traits grain yield (GRYLD) and thousand kernel weight (TKW); Ludhiana (LDH), Jabalpur (JBP); Threshold = –log10(p–value) > 3.

10.3389/fpls.2024.1419227

TABLE 5 Genome wide significant associations (R^2) of single nucleotide polymorphisms (SNPs) with agronomic traits in wheat at Jabalpur.

Trait	SNP	Chr	Alleles	Position (Mb)	p-Value	add_effect	PVE%
DAYSMT_J	S3A_1395864	3A	C/T	1.4	6.30E-04	0.67	11
	S3D_553430673	3D	C/A	553.43	4.63E-04	0.64	11
	S7A_652922600	7A	C/G	652.92	1.41E-04	0.61	13
	SUN_37992374	UN	C/T	37.99	7.97E-04	-1.16	10
DTHD_J	S1B_511789038	1B	C/T	511.79	3.30E-05	-0.38	15
	S4B_525119244	4B	G/A	525.12	8.35E-05	0.99	14
	SUN_37992374	UN	C/T	37.99	7.65E-05	-2.48	14
GRYLD_J	S1A_10214692	1A	T/A	10.21	1.42E-04	0.04	13
	S1A_3613616	1A	T/C	3.61	6.04E-04	0.02	11
	S1B_460449239	1B	G/A	460.45	6.39E-04	0.01	11
	S3A_531631224	3A	C/T	531.63	4.48E-04	0.09	11
	S3B_584821928	3B	G/T	584.82	9.19E-04	0.08	10
	S4B_648759658	4B	A/G	648.76	4.95E-04	0.05	11
	S5A_556823005	5A	G/A	556.82	2.50E-04	-0.23	12
	S5A_564814316	5A	A/G	564.81	9.41E-04	0.11	10
	S5A_564557123	5A	T/C	564.56	9.41E-04	-0.15	10
	S7A_588008821	7A	A/G	588.01	4.84E-04	-0.04	11
	S7B_707946706	7B	G/A	707.95	6.58E-04	0.17	11
PH_J	S1A_517587620	1A	G/T	517.59	7.44E-08	0.21	26
	S2A_458243706	2A	G/A	458.24	9.68E-08	-0.19	26
	S2A_590357439	2A	C/T	590.36	1.14E-07	-0.11	25
	S2A_276745579	2A	C/T	276.75	1.20E-07	0.05	25
	S2A_260549730	2A	A/G	260.55	1.21E-07	0.02	25
	S3D_18502122	3D	T/C	18.5	3.34E-08	0.62	27
	S4D_488665508	4D	A/C	488.67	7.76E-06	-0.25	18
	S6A_493375057	6A	C/A	493.38	8.32E-06	-2.46	18
	S6B_676001811	6B	A/G	676	1.10E-07	0.12	25
	S6D_454988874	6D	A/G	454.99	1.19E-07	4.65	25
TKW_J	S2D_533272704	2D	T/C	533.27	5.65E-04	2.86	11
	S2D_560760382	2D	G/A	560.76	6.02E-04	3.1	11
	S2D_556892142	2D	T/C	556.89	7.02E-04	2.99	11
	S2D_541480094	2D	C/A	541.48	8.28E-04	3.01	10
	S2D_547614644	2D	C/T	547.61	8.28E-04	-3.01	10
	S2D_543660155	2D	A/G	543.66	9.11E-04	-2.9	10
	S3A_512561741	3A	A/G	512.56	6.69E-04	-1.71	11
	S4A_688796573	4A	T/C	688.8	4.10E-04	0.54	11
	S4A_688406385	4A	G/C	688.41	6.09E-04	0.65	11
	S4A_679160910	4A	C/G	679.16	6.28E-04	-0.37	11

(Continued)

TABLE 5 Continued

Trait	SNP	Chr	Alleles	Position (Mb)	p-Value	add_effect	PVE%
	S4A_688406519	4A	C/T	688.41	7.54E-04	-0.6	10
	S6B_393496773	6B	T/C	393.5	2.50E-04	-0.36	12
	S7A_24940330	7A	G/A	24.94	1.78E-04	-1.27	13
	S7A_24923657	7A	G/A	24.92	5.95E-04	-1.45	11
	S7A_26061799	7A	A/G	26.06	7.42E-04	1.72	11
	S7B_729805573	7B	G/A	729.81	9.99E-04	-1.27	10

DTHD, days to heading; DAYSMT, days to maturity; GRYLD, grain yield; TKW, thousand-kernel weight, percent variance explained (PVE).

TABLE 6 A list of predicted proteins and function translated by IWGSC genes overlapping 19 Novel SNPs.

SNP_Position	Traits related	IWGSC ID	Predicted protein	Function	References
S1A_49153543	SDSS	TraesCS1A02G066900	SPD1	The SPD1 gene encodes a member of the AAA+ ATPase superfamily involved in plastid development during early seedling growth.	Ruppel et al., 2011
S1A_49281757	SDSS	TraesCS1A02G067300	RRP5	All results support an involvement of the analyzed proteins in ribosome biogenesis but differences in rRNA processing, gametophyte and embryo development suggested an alternative regulation in plants	Missbach et al., 2013
S2B_477569164	SDSS	TraesCS2B02G333900	Phytosiderophore biosynthesis like DMAS1-B, NRAMP2 and NAAT2-D	Fe/Zn transport and accumulation in grain	Gupta et al., 2020
S2B_753091778	Twt	TraesCS2B02G559000	B30.2/SPRY_sf	The B30.2/SPRY domain in these proteins is likely to function through protein-protein interaction	Woo et al., 2006
S2D_69502623	PH	TraesCS2D02G120100	GEX1/Brambleberry	GEX1 from Arabidopsis is required for correct pollen maturation	Alandete-Saez et al., 2011
S3A_512561741	TKW	TraesCS3A02G284100	Galacturonosyltransferase (GAUT)	GAUTs are involved in pectin and xylan biosynthesis	Bouton et al., 2002
S3A_531631224	GYLD	TraesCS3A02G296900	PyrdxlP-dep_Trfase_dom1	PLP-dependent enzymes are primarily involved in the biosynthesis of amino acids and amino acid-derived metabolites, but they are also found in the biosynthetic pathways of amino sugars and in the synthesis or catabolism of neurotransmitters; pyridoxal phosphate can also inhibit DNA polymerases and several steroid receptors	Mozzarelli and Bettati., 2006
S3B_720255460	P%	TraesCS3B02G471800	LRR_dom_sf/NB-ARC	Proteins containing LRRs include tyrosine kinase receptors, cell-adhesion molecules, virulence factors, and extracellular matrix-binding glycoproteins, and are involved in a variety of biological processes, including signal transduction, cell adhesion, DNA repair, recombination, transcription, RNA processing, disease resistance, apoptosis, and the immune response	van der Biezen and Jones., 1998
S3B_728890092	P%	TraesCS3B02G481200	F-box-like_dom_sf/F-box_dom	First identified in cyclin-F as a protein-protein interaction motif, the F-box is a conserved domain that is present in numerous proteins with a bipartite structure	Bai et al., 1996
S3B_739166411	TKW	TraesCS3B02G494600	Conserved oligomeric Golgi complex, subunit 4 (COG_su4)	COG4 is a component of the conserved oligomeric Golgi (COG) complex which mediates the proper glycosylation of proteins trafficking through the Golgi apparatus. It is included in the CATCHR	Santana-Molina et al., 2021

(Continued)

TABLE 6 Continued

SNP_Position	Traits related	IWGSC ID	Predicted protein	Function	References
				(complexes associated with tethering containing helical rods) family, which includes components of the exocyst, GARP, and DSL1 complexes and share structural and functional features: the α-helical bundles at the middle/C-terminal (described as domains A-D/E) and a N-terminal coiled-coil region.	
S3B_768723701	Twt	TraesCS3B02G526500	Bax_inhibitor_1-related	BI-1 also regulates cell death triggered by ER stress. BI-1 appears to exert its effect through an interaction with calmodulin	Weis et al., 2013
S4A_679160910	TKW	TraesCS4A02G406300	DHNA_phytyltransferase_MenA	2-carboxy-1,4-naphthoquinone phytyltransferase (IPR011937)	Johnson et al., 2000
S4A_740926925	SDSS	TraesCS4A02G491100	Thioredoxin-like_sf	Several biological processes regulate the activity of target proteins through changes in the redox state of thiol groups (S2 to SH2), where a hydrogen donor is linked to an intermediary disulphide protein. Such processes include the ferredoxin/thioredoxin system, the NADP/thioredoxin system, and the glutathione/glutaredoxin system. Several of these disulphide proteins share a common structure, consisting of a three-layer α/β/α core. Proteins that contain domains with a thioredoxin-like fold	Buchanan and Balmer., 2005
S4D_488665508	PH	TraesCS4D02G330500	Helix-loop-helix DNA-binding domain superfamily (HLH_DNA-bd_sf)	A number of eukaryotic proteins, which probably are sequence specific DNA-binding proteins that act as transcription factors, share a conserved domain of 40 to 50 amino acid residues. The proteins of this subfamily act together with co-repressor proteins, like groucho, through their -terminal motif WRPW.	Murre et al., 1989
S5A_556823005	GYLD	TraesCS5A02G354200	Not available	Not available	Not available
S7A_13179057	P%	TraesCS7A02G031700	Aminoacyl-tRNA synthetase, class II (D/K/N) (IPR004364)	The aminoacyl-tRNA synthetases (also known as aminoacyl-tRNA ligases) catalyze the attachment of an amino acid to its cognate transfer RNA molecule in a highly specific two-step reaction. These proteins differ widely in size and oligomeric state, and have limited sequence homology.	Woese et al., 2000
S7A_652922600	DAYSMT	TraesCS7A02G458100	Znf_RING/FYVE/PHD	Znf-containing proteins function in gene transcription, translation, mRNA trafficking, cytoskeleton organization, epithelial development, cell adhesion, protein folding, chromatin remodeling and zinc sensing, to name but a few	Matthews and Sunde., 2002
S7B_689968561	GH	TraesCS7B02G420600	Peptidase S28/Alpha/Beta hydrolase fold	Serine carboxypeptidase S28 family comprises carboxypeptidase PRCP and the aminopeptidase DPP7. The cap domain (SKS) is formed by 11 α-helices and two strands interconnected by loops. It contains four disulphide bonds which are assumed to be involved in stabilizing the structure. The SKS domain is a rare fold possibly present only in the S28 serine peptidase family.	Bezerra et al., 2012
S7D_476139586	TKW	TraesCS7D02G367800	Major intrinsic protein (IPR000425)	The major intrinsic protein (MIP) family is large and diverse, possessing over 100 members that form transmembrane channels. These channel proteins function in water, small carbohydrate (e.g., glycerol), urea, NH3, CO2 and possibly ion transport, by an energy independent mechanism.	Fu et al., 2000

Grain protein is the primary determinant of wheat quality, its end use, and commercial value (Cox et al., 1985). However, it is well known that grain protein is negatively correlated with grain yield in wheat; in our study, too, we found that there was a negative correlation between protein percentage and the yield from Jabalpur (-0.20), but there was a slight non-significant correlation between protein % and the yield from Ludhiana (0.11). This revealed the influence of environment on the genotype quality.

Despite this negative correlation, many reports have reported simultaneous improvements in grain yield and GP (Niu et al., 2010; Vishwakarma et al., 2014, 2016). Protein % has a positive correlation with all the quality measures. Even in this study, we found that all the quality traits studied significantly correlated. Mladenov et al. (2012) also reported a significant positive correlation between the Protein % and sedimentation (SDSS value). In addition to this, test weight is a measure of grain density, which showed a significantly positive correlation with Protein %; in contrast, none of the other traits had a significant correlation. This elucidates that the previous study's test weight had shown a positive correlation with TKW but not with the Protein % and SDSS value due to these being environment-specific (Mladenov et al., 2012). Grain yield showed a positive correlation (0.17*) with days to maturity indicating that an increase in days to maturity would increase grain yield as also mentioned by Semnaninejad et al. (2021). In addition to this, Grain yield significantly positively correlated with TKW_L (0.39***) and TKW_J (0.04) for both locations as identified by earlier reports (Semnaninejad et al., 2021; Zhang et al., 2021).

GWAS for agronomic traits

For days to heading, we identified two MTAs on 4D. For the Jabalpur location, it was on chromosome 1B, 4B UN, which was earlier reported by Cadalen et al. (1998) in a double haploid population with his model that explained that the heading date loci from chromosomes 4B and 4D (Xfba1- 4B, Xglk556–4B, and Xfba211–4D) had the main effects. There were interaction effects with plant height QTLs (Xfba393–1A and Xcdo1188–1B) which explained about 50% of the plant height variation. Worland (1996) elucidated that almost all chromosomes carry genes for heading. This notwithstanding, the important genes *Vrn* (vernalization) and *Ppd* (photoperiod), located in homeologous groups 5 and 2, have a significant role in heading date. The SNPs significantly associated with plant height were identified on 2B, 2 (2D), 7B, and UN for the Ludhiana location and 1A, 2A, 3D, 4D, 6A, 6B, and 6D for the Jabalpur location. Previously, plant height was also reported on chromosomes 1A (Sukumaran et al., 2015), 2A (Ain et al., 2015; Mengistu et al., 2016; Sheoran et al., 2019), and 2B (Zanke et al., 2014; Ain et al., 2015; Gao et al., 2015; Sheoran et al., 2019). In addition, the SNPs identified for PH on chromosome 2B (565.060 Mb) were found in proximity to the reduced height genes Rht4 (609.3 Mb). We also obtained 4 SNPs on chromosome 2A, where the Rht7 gene was reported. In this study, SNPs for PH were detected on chromosome 6A. The locus on 6A was consistently detected under drought, heat, and irrigated conditions for yield (Edae et al., 2015; Lopes et al., 2015). Sukumaran et al. (2015) reported the PH in the WAMI population and PH was not correlated with YLD according to the genetic and phenotypic correlation study, demonstrating that the loci on 6A have pleiotropic effects on several characteristics (Sukumaran et al., 2015). However, numerous studies have shown that QTLs influenced by environmental factors in various crops regulate plant height and heading date (Xu et al., 2005; Zhang et al., 2009). The SNPs significantly associated with DTM were identified on chromosomes 4A, 4B, and 5A for the Ludhiana location and 3A, 3D, and 7A for the Jabalpur location, corresponding to the earlier reported genomic regions for DTM on chromosome 5A (Gahlaut et al., 2019; Sheoran et al., 2019), 4B (Sukumaran et al., 2015), and 7A (Adhikari et al., 2020). SNP S7A_652922600 (*TraesCS7A02G458100*) for the DAYSMT on chromosome 7A plays a key role in the function of Znf-containing proteins.

We found a set of three markers for TKW on chromosome 7A in the region from 718 to 735 Mb while Rathan et al. (2022) and Jamil et al. (2019) also identified markers for TKW on the same chromosomal region at 731.8 Mb in multiple environments. This indicated that this chromosome region might have some haplotype block for the TKW.

A complicated quantitative feature, grain yield contains MTAs dispersed across several chromosomes (Jamil et al., 2019). For grain yield in our study, we found MTAs on chromosomes 1A and 4D for the Ludhiana location, while for the Jabalpur location there were three MTAs on chromosome 5A and two MTAs on 1A, 1B, and 3A. Jamil et al. (2019) reported QTLs on 1A, 1B, 5A, and 3A for GRYLD. In an earlier study, two MTAs were present on chromosome 1A with 13% PVE. While chromosome 1B QTLs had 7.55% PVE in our study, one common MTA (S1A_3613616) on 1A had a pleiotropic effect with grain hardness that had a positive correlation with GRYLD also, which intimated that GH had a direct effect on GRYLD. Three markers on 1B explained 13% to 16% PVE for GRYLD (Jamil et al., 2019). Moreover, we reported one MTA on the 4D chromosome. Li et al. (2014) also reported a QTL (QGy4D) on the same chromosome flanked by SSR marker Xbarc334-Xwmc331. It is recognized that several important genes regulating plant height, yield productivity, and yield components are located on chromosomes 4B and 4D (Huang et al., 2006). In our study, we used different environments with very distinct climates. Jabalpur has high ambient temperatures in the daytime and cooler nights during the crop growing season, and the crop's days to heading, flowering, and maturity periods are shorter in comparison to Ludhiana. In contrast, Ludhiana has a significantly colder environment with longer growing seasons (Mondal et al., 2016). These differences in the environment were also seen in the marker-trait association for the agronomic traits at both locations. We assume that this was the main reason for there being no common MTAs identified for the agronomic traits.

GWAS for quality and values added parameters

For the quality traits, most of the genetics studies undertaken on wheat have used linkage mapping to study the genetic basis of quality determinants. This entails identifying genes/QTLs linked with the trait of interest by establishing linkage disequilibrium (LD) in populations obtained from bi-parental crosses. However, because of the limited number and location of meiotic events, QTL mapping resolution is frequently limited to 10–30 cM, and it can only study a small fraction of the total number of potential alleles in the

population from which the parents originated (Zhu et al., 2008). As an alternative to linkage mapping, association mapping (AM) can help locate alleles in a large number of germplasm samples Yu et al., 2006). Earlier investigations revealed that the GLM model could create false-positive sites due to the lack of a Kinship matrix and a shift in the phenotypic interpretation rate (Yu et al., 2006).

Value-added parameters are complex traits influenced by both the genetic background of the germplasm and the growth conditions (Mohan et al., 2022). Earlier genes/QTLs with major and minor effects on wheat end-use quality traits have been identified and characterized. Nonetheless, whether with bi-parental or association studies, genetically dissected the quantitative trait loci (QTLs)/alleles for the GA Score, PR Score, and SDSS Value, these traits remain uncharacterized. The literature search revealed that this is the first time MTAs for GA Score (2), PR score (5), and the SDSS Value (13) have been reported. This novel locus can be helpful in the identification of new end-use quality products with their corresponding combination in the existing germplasm.

The SDSS value is a thorough indicator for subtly assessing wheat quality and one of the crucial tests to gauge flour's gluten content. This directly affects the flour's suitability for processing and baking (Peña et al., 2012) Given that the SDSS value is a quantitative variable influenced by genetic and environmental influences, some QTLs can only be found in particular environments. We reported 13 such MTAs, which are consistent with other findings located on chromosomes 1A (6 SNPs), 1B (2), 2B (3), 4A, and 7B (Goel et al., 2019; Zhang et al., 2020; Alemu et al., 2021). We reported SNP S1A_510849238 on 1A, were near to QTL (540,660,000–544,610,000 bp, RefSeqv1.0) as earlier reported by Yang et al. (2020) through a GWAS. In addition, two more QTLs were reported on chromosome 1A QSsv.cau-1A.1.1 (371,573,909–386,426,688 bp, RefSeqv1.0) and QSsv.cau-1A.1.2 (419,490,584–492,004,197 bp, RefSeqv1.0) by (Tian et al., 2021). We found six MTAs associated with SDSS value on chromosome 1A; this indicates that chromosome 1A is an important region for SDSS value.

According to research, grains that react with phenol to generate color also have the unfavorable trait of browning wheat products like pasta and noodles (Bernier and Howes, 1994). This makes grain screening a valuable method for determining the quality of the end product, thus proving useful in screening the end products. For the phenol reaction score, we have reported five MTAz in between 712.85 -706.42 Mb i.e., the 6.43 Mb region only; this elucidated the possibility of a haplotype block for this trait. The phenol color reaction of the grain gene was on the long arm of chromosome 2A. According to Nair and Tomar (2001) *Triticum turgidum* variety *durum* Desf. has at least two genes that regulate the phenol color response.

The GA score was evaluated based on the grain's size, shape, color, and luster. We have reported only two MTAs for the GA score on 5A and 7B chromosomes. To date, no previous report has been found for the gene/QTLs for GA score. Previously, Kumar et al. (2019) measured six traits related to grain shape and size, namely, length, width, area, length-to-width ratio, test weight, and thousand kernel weight. Despite a significant correlation with grain yield traits, no significant QTL was found for these traits. These findings could lead to the hypothesis that focusing on grain shape and size, particularly an increase in GA, may improve wheat yield by increasing TGW. Test weight is often referred to as the specific weight of a known volume of grain and serves as a crucial quality indicator. We reported MTAs for test weight on chromosomes 1A, 2A (3), 2B, 3A, 3B, 6D (2), 7A (2), and 7B. There are few studies showing QTL for test weight; however, one of the most recent ones found eight loci on chromosomes 1D, 2A, 2B, 2D, 3B, 3D, 4D, and 7A (Cabral et al., 2018), and while another found loci on 1B and 3B (Alemu et al., 2021).

The absence of Gpc-B1 allows the exploration of the novel identified loci contributed by the lines

Our study did not identify the major Gpc-B1 gene reported on chromosome 6B by Uauy et al. (2006). This indicates that the genotype x environment interactions played a crucial role. Therefore, exploring a non-adapted genotype provides an opportunity to enhance GPC in the cultivated wheat gene pool. *Gpc-B1* played a significant role in developing several lines for the grain protein; however, it was found at par or negative yield (Uauy et al., 2006; Blanco et al., 2012; Vishwakarma et al., 2014, 2016). These independent loci could be useful to enhance GPC through MAS, without compromising yield. In this study, we identified MTAs on chromosomes 1A, 1B, 3B, 5B, and 7A. In earlier reports with bi-parental mapping populations, *QGPC.ndsu.5B* (found on 5BS) and *QGPC.ndsu.7A.2* (found on 7AL) QTLs were present in non-adapted germplasm, according to a comparison with 49 GPC investigations (Kumar A. et al., 2018). Even though El-Feki et al. (2013) discovered a QTL on 5BS, it was too far away from *QGPC.ndsu.5B*. A few previous investigations in both durum (Peleg et al., 2009; Suprayogi et al., 2009) and hexaploid wheat (Mann et al., 2009; Li et al., 2012) found a stable QTL for GPC on 7AL. The QTL *QGPC.ndsu.7A.2* was found near the telomeric end of chromosomal arm 7AL, whereas the QTLs previously published (Mann et al., 2009; Peleg et al., 2009; Suprayogi et al., 2009; Li et al., 2012) were found in the middle of the chromosome arm 7AL. We reported five MTAs on 3B alone within the 20.98 Mb region. The presence of GPC region *QGpc.caas-3B* flanked by marker wmc3-wmc418in in bi-parental mapping that showed a high LOD value, 11.10, with the highest phenotypic variance of 14.5% has been reported previously by Li et al. (2009). This could be of significant interest as these QTLs were independent of grain yield and may be used as haplotype blocks, contributing to the favorable alleles in the future.

Novel allele for grain hardness on chromosome 7B

Grain hardness or texture in wheat is directly associated with critical end-use quality attributes such as milling yield and flour extraction. Our research corroborated this by indicating a moderately positive relationship between GH and flour extraction in all settings. Grain hardness in wheat is controlled by the main

hardness locus (Ha) on chromosome 5DS, which is positioned at a sub-telomeric location (Sourdille et al., 1996; Morris, 2002). Friabilins are 15-kD lipid-binding endosperm-specific proteins encoded by the Ha locus. The two main proteins in friabilins are Puroindoline a (Pina) and Puroindoline b (Pinb) (Gautier et al., 1994). According to a study of diverse wheat sets (Morris, 2002), hard wheat varieties either lack or possess specific mutations for the pin coding genes. The wild-type pin alleles are found in soft wheat types (Bhave and Morris, 2008). In addition to the significance of the Ha gene, previous research has identified numerous additional QTLs linked to hardness (Heo and Sherman, 2013). In this study, we found seven MTAs on 7B, and no QTL for GH on 7B was reported. These MTAs on 7B could be novel alleles, indicating that both parental genotypes will likely contain the Ha locus hardness alleles.

Conclusion

This is the first study to report a GWAS for value added quality traits in bread wheat *T. aestivum*. Genetic and functional analysis of the associated genomic regions may enhance wheat quality. Overall, several lines with a combination of appropriate grain quality and agronomic traits were identified, especially for protein content that plays a vital role in tackling nutritional deficiencies or hidden hunger. Quality-enriched *T. aestivum* lines and genomic regions harboring grain quality SNPs can accelerate the breeding program for developing nutritional and value-added end product quality wheat 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

MV: Formal analysis, Investigation, Writing – original draft, Writing – review & editing. PB: Data curation, Writing – review & editing. UK: Conceptualization, Project administration, Writing – review & editing. RS: Conceptualization, Funding acquisition, Supervision, Writing – review & editing. SK: Visualization, Writing – review & editing. VG: Resources, Visualization, Writing – review & editing. GM: Data curation, Formal analysis, Resources, Writing – review & editing. KT: Data curation, Writing – review & editing. ND: Writing – review & editing, Data curation, Formal analysis. AJ: Resources, Visualization, Conceptualization, Funding acquisition, Project administration, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. The financial support was provided by the US Agency for International Development (USAID) Feed the Future Innovation Lab for Applied Wheat Genomics (Cooperative Agreement No. AID-OAA-A-13–00051) and Department of Biotechnology, Govt. of India (Grant No. BT/Ag/Network/Wheat/2019–20).

Acknowledgments

The technical support provided by Mr. Avdhesh Kumar and Dr. Raj Kumar Jat is duly acknowledged.

Conflict of interest

Author UK was employed by the company Astralyan Agro OPC Pvt. Ltd.

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2024.1419227/full#supplementary-material

References

AACC (2000). *Approved Methods of the American Association*.

Adhikari, A., Basnet, B. R., Crossa, J., Dreisigacker, S., Camarillo, F., Bhati, P. K., et al. (2020). Genome-wide association mapping and genomic prediction of anther extrusion in CIMMYT hybrid wheat breeding program via modeling pedigree, genomic relationship, and interaction with the environment. *Front. Genet.* 11, 586687. doi: 10.3389/fgene.2020.586687

Ain, Q. U., Rasheed, A., Anwar, A., Mahmood, T., Imtiaz, M., Mahmood, T., et al. (2015). Genome-wide association for grain yield under rainfed conditions in historical wheat cultivars from Pakistan. *Front. Plant Sci.* 6, 743. doi: 10.3389/fpls.2015.00743

Alandete-Saez, M., Ron, M., Leiboff, S., and McCormick, S. (2011). Arabidopsis thaliana GEX1 has dual functions in gametophyte development and early embryogenesis. *Plant J. Cell Mol. Biol.* 68 (4), 620–632. doi: 10.1111/j.1365-313X.2011.04713.x

Alemu, A., El Baouchi, A., El Hanafi, S., Kehel, Z., Eddakhir, K., Tadesse, W., et al. (2021). Genetic analysis of grain protein content and dough quality traits in elite spring bread wheat (*Triticum aestivum*) lines through association study. *J. Cereal Sci.* 100, 103214. doi: 10.1016/j.jcs.2021.103214

Alvarado, G., Rodríguez, F. M., Pacheco, A., Burgueño, J., Crossa, J., Vargas, M., et al. (2020). META-R: A software to analyze data from multi-environment plant breeding trials. *Crop J.* 8, 745–756. doi: 10.1016/j.cj.2020.03.010

Bai, C., Sen, P., Hofmann, K., Ma, L., Goebl, M., Harper, J. W., et al. (1996). SKP1 connects cell cycle regulators to the ubiquitin proteolysis machinery through a novel motif, the f-box. *Cell* 86 (2), 263–274. doi: 10.1016/s0092-8674(00)80098-7

Bernier, A. M., and Howes, N. K. (1994). Quantification of variation in tyrosinase activity among durum and common wheat cultivars. *J. Cereal Sci.* 19, 157–159. doi: 10.1006/jcrs.1994.1021

Bezerra, G. A., Dobrovetsky, E., Dong, A., Seitova, A., Crombett, L., Shewchuk, L. M., et al. (2012). Structures of human DPP7 reveal the molecular basis of specific inhibition and the architectural diversity of proline-specific peptidases. *PloS One* 7 (8), e43019. doi: 10.1371/journal.pone.0043019

Bhatta, M., Morgounov, A., Belamkar, V., and Baenziger, P. S. (2018). Genome-wide association study reveals novel genomic regions for grain yield and yield-related traits in drought-stressed synthetic hexaploid wheat. *Int. J. Mol. Sci.* 19, 3011. doi: 10.3390/ijms19103011

Bhave, M., and Morris, C. F. (2008). Molecular genetics of puroindolines and related genes: Allelic diversity in wheat and other grasses. *Plant Mol. Biol.* 66, 205–219. doi: 10.3390/ijms19103011

Blanco, A., Mangini, G., Giancaspro, A., Giove, S., Colasuonno, P., Simeone, R., et al. (2012). Relationships between grain protein content and grain yield components through quantitative trait locus analyses in a recombinant inbred line population derived from two elite durum wheat cultivars. *Mol. Breeding* 30, 79–92. doi: 10.1007/s11032-011-9600-z

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

Bouton, S., Leboeuf, E., Mouille, G., Leydecker, M. T., Talbotec, J., Granier, F., et al. (2002). QUASIMODO1 encodes a putative membrane-bound glycosyltransferase required for normal pectin synthesis and cell adhesion in arabidopsis. *Plant Cell* 14 (10), 2577–2590. doi: 10.1105/tpc.004259

Buchanan, B. B., and Balmer, Y. (2005). Redox regulation: a broadening horizon. *Annu Rev Plant Biol.* 56, 187–220. doi: 10.1146/annurev.arplant.56.032604.144246

Cabral, A. L., Jordan, M. C., Larson, G., Somers, D. J., Humphreys, G., and McCartney, C. A. (2018). Relationship between QTL for grain shape, grain weight, test weight, milling yield, and plant height in the spring wheat cross RL4452/ 'AC Domain.'. *PloS One* 13, e0190681. doi: 10.1371/journal.pone.0190681

Cadalen, T., Sourdille, P., Charmet, G., Tixier, M. H., Gay, G., Boeuf, C., et al. (1998). Molecular markers linked to genes affecting plant height in wheat using a doubled-haploid population. *Theor. Appl. Genet.* 96, 933–940. doi: 10.1007/s001220050823

Cappelli, A., and Cini, E. (2021). Challenges and opportunities in wheat flour, pasta, bread, and bakery product production chains: A systematic review of innovations and improvement strategies to increase sustainability, productivity, and product quality. *Sustainability* 13, 2608. doi: 10.3390/su13052608

Carter, B. P., Morris, C. F., and Anderson, J. A. (1999). Optimizing the SDS sedimentation test for end-use quality selection in a soft white and club wheat breeding program. *Cereal Chem.* 76, 907–911. doi: 10.1094/CCHEM.1999.76.6.907

Cox, M. C., Qualset, C. O., and Rains, D. W. (1985). Genetic variation for nitrogen assimilation and translocation in wheat. I. Dry matter and nitrogen accumulation. *Crop Sci.* 25, 430–435. doi: 10.2135/cropsci1985.0011183X002500030002x

Doyle, J. J., and Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* 19, 11–15.

Edae, E. A., Bowden, R. L., and Poland, J. (2015). Application of population sequencing (POPSEQ) for ordering and imputing genotyping-by-sequencing markers in hexaploid wheat. *G3 Genes Genomes Genet.* 5, 2547–2553. doi: 10.1534/g3.115.020362

El-Feki, W. M., Byrne, P. F., Reid, S. D., Lapitan, N. L. V., and Haley, S. D. (2013). Quantitative trait locus mapping for end-use quality traits in hard winter wheat under contrasting soil moisture levels. *Crop Sci.* 53, 1953–1967. doi: 10.2135/cropsci2012.12.0674

Fu, D., Libson, A., Miercke, L. J., Weitzman, C., Nollert, P., Krucinski, J., et al. (2000). Structure of a glycerol-conducting channel and the basis for its selectivity. *Sci. (New York N.Y.)* 290 (5491), 481–486. doi: 10.1126/science.290.5491.481

Gahlaut, V., Jaiswal, V., Tyagi, B. S., Singh, G., Sareen, S., Balyan, H. S., et al. (2019). Multi-locus genome wide association mapping for yield and its contributing traits in hexaploid wheat under different water regimes. *Sci. Rep.* 9, 19486. doi: 10.1038/s41598-019-55520-0

Gao, F., Wen, W., Liu, J., Rasheed, A., Yin, G., Xia, X., et al. (2015). Genome-wide linkage mapping of QTL for yield components, plant height and yield-related physiological traits in the Chinese wheat cross Zhou 8425B/Chinese spring. *Front. Plant Sci.* 6, 1099. doi: 10.3389/fpls.2015.01099

Gautier, M. F., Aleman, M. E., Guirao, A., Marion, D., and Joudrier, P. (1994). *Triticum aestivum* puroindolines, two basic cystine-rich seed proteins: cDNA sequence analysis and developmental gene expression. *Plant Mol. Biol.* 25, 43–57. doi: 10.1007/BF00024197

Glaubitz, J. C., Casstevens, T. M., Lu, F., et al. (2014). TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. *PloS One* 9, e90346. doi: 10.1371/journal.pone.0090346

Goel, S., Singh, K., Singh, B., Grewal, S., Dwivedi, N., Alqarawi, A. A., et al. (2019). Analysis of genetic control and QTL mapping of essential wheat grain quality traits in a recombinant inbred population. *PloS One* 14, e0200669. doi: 10.1371/journal.pone.0200669

Güçbilmez, Ç. M., Şahin, M., Akçacık, A. G., Aydoğan, S., Demir, B., Hamzaoğlu, S., et al. (2019). Evaluation of GlutoPeak test for prediction of bread wheat flour quality, rheological properties and baking performance. *J. Cereal Sci.* 90, 102827. doi: 10.1016/j.jcs.2019.102827

Gupta, O. P., Pandey, V., Saini, R., Narwal, S., Malik, V.K., Khandale, T., et al. (2020). Identifying transcripts associated with efficient transport and accumulation of Fe and Zn in hexaploid wheat (*T. aestivum* L.). *J. Biotechnol.* 316, 46–55. doi: 10.1016/j.jbiotec.2020.03.015

Guzmán, C., Ammar, K., Govindan, V., and Singh, R. (2019). Genetic improvement of wheat grain quality at CIMMYT. *Front. Agric. Sci. Eng.* 6, 265–272. doi: 10.15302/J-FASE-2019260

Guzmán, C., Peña, R. J., Singh, R., Atrique, E., Dreisigacker, S., Crossa, J., et al. (2016). Wheat quality improvement at CIMMYT and the use of genomic selection on it. *Appl. Transl. Genom.* 11, 3–8. doi: 10.1016/j.atg.2016.10.004

Hassani-Pak, K., Singh, A., Brandizi, M., Hearnshaw, J., Parsons, J.D., Amberkar, S., et al. (2021). KnetMiner: a comprehensive approach for supporting evidence-based gene discovery and complex trait analysis across species. *Plant Biotechnol. J.* 19, 1670–1678. doi: 10.1111/pbi.13583

Heo, H., and Sherman, J. (2013). Identification of QTL for grain protein content and grain hardness from winter wheat for genetic improvement of spring wheat. *Plant Breed Biotechnol.* 1, 347–353. doi: 10.9787/PBB.2013.1.4.347

Huang, X. Q., Cloutier, S., Lycar, L., et al. (2006). Molecular detection of QTLs for agronomic and quality traits in a doubled haploid population derived from two Canadian wheats (*Triticum aestivum* L.). *Theor. Appl. Genet.* 113, 753–766. doi: 10.1007/s00122-006-0346-7

Ibba, M. I., Crossa, J., Montesinos-López, O. A., et al. (2020). Genome-based prediction of multiple wheat quality traits in multiple years. *Plant Genome* 13, e20034. doi: 10.1002/tpg2.20034

IWGSC (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* 80-). 361, 1–163.

Jamil, M., Ali, A., Gul, A., Ghafoor, A., Napar, A. A., Ibrahim, A. M. H., et al. (2019). Genome-wide association studies of seven agronomic traits under two sowing conditions in bread wheat. *BMC Plant Biol.* 19, 149. doi: 10.1186/s12870-019-1754-6

Jiang, Y., Schmidt, R. H., Zhao, Y., and Reif, J. C. (2017). Quantitative genetic framework highlights the role of epistatic effects for grain-yield heterosis in bread wheat. *Nat. Genet.* 49, 1741–1746. doi: 10.1038/ng.3974

Johnson, T. W., Shen, G., Zybailov, B., Kolling, D., Reategui, R., Beauparlant, S., et al. (2000). Recruitment of a foreign quinone into the A(1) site of photosystem i. i. genetic and physiological characterization of phylloquinone biosynthetic pathway mutants in synechocystis sp. pcc 6803. *J. Biol. Chem.* 275 (12), 8523–8530.

Juliana, P., Poland, J., Huerta-Espino, J., Shrestha, S., Crossa, J., Crespo-Herrera, L., et al. (2019). Improving grain yield, stress resilience and quality of bread wheat using large-scale genomics. *Nat. Genet.* 51, 1530–1539. doi: 10.1038/s41588-019-0496-6

Kumar, A., Jain, S., Elias, E. M., Ibrahim, M., and Sharma, L. K. (2018). "An overview of QTL identification and marker-assisted selection for grain protein content in wheat," in *Eco-friendly Agro-biological Techniques for Enhancing Crop Productivity*. Eds. R. Sengar and A. Singh (Springer, Singapore). doi: 10.1007/978-981-10-6934-5_11

Kumar, A., Mantovani, E. E., Simsek, S., Jain, S., Elias, E. M., and Mergoum, M.. (2019). Genome wide genetic dissection of wheat quality and yield related traits and their relationship with grain shape and size traits in an elite × non-adapted bread wheat cross. *PloS One* 14, e0221826. doi: 10.1371/journal.pone.0221826

Kumar, S., Sohu, V. S., Gupta, S. K., Singh, R. P., and Bains, N. S. (2018). Understanding the chapati making attributes of the Indian wheats – I: The physico-chemical basis. *J. Appl. Natural Sci.* 10, 572–592. doi: 10.31018/jans.v10i2.1739

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923

Li, J., Cui, F., Ding, Am., Zhao, C-h., Wang, X-q., Wang, L., et al. (2012). QTL detection of seven quality traits in wheat using two related recombinant inbred line populations. *Euphytica* 183, 207–226. doi: 10.1007/s10681-011-0448-4

Li, Z. K., Jiang, X. L., Peng, T., Shi, C. L., Han, S. X., Tian, B., et al. (2014). Mapping quantitative trait loci with additive effects and additive x additive epistatic interactions for biomass yield, grain yield, and straw yield using a doubled haploid population of wheat (*Triticum aestivum* L.). *Genet. Mol. Res.* 13, 1412–1424. doi: 10.4238/2014.February.28.14

Li, Y., Song, Y., Zhou, R., Branlard, G., and Jia, J. (2009). Detection of QTLs for bread-making quality in wheat using a recombinant inbred line population. *Plant Breeding* 128, 235–243. doi: 10.1111/j.1439-0523.2008.01578.x

Liu, Y., Lin, Y., Gao, S., Li, Z., Ma, J., Deng, M., et al (2017). A genome-wide association study of 23 agronomic traits in Chinese wheat landraces. *Plant J.* 91, 861–873. doi: 10.1111/tpj.13614

Lopes, M. S., Dreisigacker, S., Peña, R. J., Sukumaran, S., and Reynolds, M. P. (2015). Genetic characterization of the wheat association mapping initiative (WAMI) panel for dissection of complex traits in spring wheat. *Theor. Appl. Genet.* 128, 453–464. doi: 10.1007/s00122-014-2444-2

Ma, F., and Baik, B. K. (2018). Soft wheat quality characteristics required for making baking powder biscuits. *J. Cereal Sci.* 79, 127–133. doi: 10.1016/j.jcs.2017.10.016

Mann, G., Diffey, S., Cullis, B., Azanza, F., Martin, D., Kelly, A., et al. (2009). Genetic control of wheat quality: Interactions between chromosomal regions determining protein content and composition, dough rheology, and sponge and dough baking properties. *Theor. Appl. Genet.* 118, 1519–1537. doi: 10.1007/s00122-009-1000-y

Maphosa, L., Langridge, P., Taylor, H., Parent, B., Emebiri, L. C., Kuchel, H., et al. (2014). Genetic control of grain yield and grain physical characteristics in a bread wheat population grown under a range of environmental conditions. *Theor. Appl. Genet.* 127, 1607–1624. doi: 10.1007/s00122-014-2322-y

Matthews, J. M., and Sunde, M. (2002). Zinc fingers–folds for many occasions. *IUBMB Life* 54 (6), 351–355. doi: 10.1080/15216540216035

Mengistu, D. K., Kidane, Y. G., Catellani, M., Frascaroli, E., Fadda, C., Pè, M. E., et al. (2016). High-density molecular characterization and association mapping in Ethiopian durum wheat landraces reveals high diversity and potential for wheat breeding. *Plant Biotechnol. J.* 14, 1800–1812. doi: 10.1111/pbi.12538

Missbach, S., Weis, B. L., Martin, R., Simm, S., Bohnsack, M. T., and Schleiff, E. (2013). 40S ribosome biogenesis co-factors are essential for gametophyte and embryo development. *PloS One* 8 (1), e54084. doi: 10.1371/journal.pone.0054084

Mladenov, V., Banjac, B., Krishna, A., and Miloaević, M. (2012). Relation of grain protein content and some agronomic traits in European cultivars of winter wheat. *Cereal Res. Commun.* 40, 532–541. doi: 10.1556/CRC.40.2012.0004

Mohan, D., Sendhil, R., Gupta, O. P., Pandey, V., Krishnappa, G., and Singh, G. P. (2022). Wheat quality index: new holistic approach to identify quality superior genotypes. *Cereal Res. Commun.* 50, 1105–1115. doi: 10.1007/s42976-022-00254-5

Mondal, S., Rutkoski, J. E., Velu, G., Singh, P. K., Crespo-Herrera, L. A., Guzmán, C., et al. (2016). Harnessing diversity in wheat to enhance grain yield, climate resilience, disease and insect pest resistance and nutrition through conventional and modern breeding approaches. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.00991

Morris, C. F. (2002). Puroindolines: The molecular genetic basis of wheat grain hardness. *Plant Mol. Biol.* 48, 633–647. doi: 10.1023/A:1014837431178

Mozzarelli, A., and Bettati, S. (2006). Exploring the pyridoxal 5'-phosphate-dependent enzymes. *Chem. Rec. (New York N.Y.)* 6 (5), 275–287. doi: 10.1002/tcr.20094

Murre, C., McCaw, P. S., and Baltimore, D. (1989). A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins. *Cell* 56 (5), 777–783. doi: 10.1016/0092-8674(89)90682-x

Nair, S. K., and Tomar, S. M. S. (2001). Genetics of phenol colour reaction of grains and glumes in tetraploid and hexaploid wheats. *J. Genet. Breed.* 55, 369–373.

Nakamura, K., Taniguchi, Y., Taira, M., and Ito, H. (2012). Investigation of soft wheat flour quality factors associated with sponge cake sensory tenderness. *Cereal Chem.* 89, 79–83. doi: 10.1094/CCHEM-07-11-0081

Niu, N., Arief, V. N., Delacy, I. H., Lush, D., Sheppard, J., Zhang, G., et al. (2010). Genetic gain in yield and protein over two cycles of a wheat recurrent selection program. *Breed Sci.* 60, 181–186. doi: 10.1270/jsbbs.60.181

Patterson, N., Price, A. L, and Reich, D. (2006). Population structure and eigenanalysis. *PloS Genet.* 2, e190. doi: 10.1371/journal.pgen.0020190

Peleg, Z., Cakmak, I., Ozturk, L., Yazici, A., Jun, Y., Budak, H., et al. (2009). Quantitative trait loci conferring grain mineral nutrient concentrations in durum wheat × wild emmer wheat RIL population. *Theor. Appl. Genet.* 119, 353–369. doi: 10.1007/s00122-009-1044-z

Peña, R. J., Trethowan, R., Pfeiffer, W. H., and van Ginkel, M. (2012). Quality (end-use) improvement in wheat compositional, genetic, and environmental factors. *J. Crop Prod.* 5, 1–37. doi: 10.1300/j144v05n01_02

Poland, J. A., Brown, P. J., Sorrells, M. E., and Jannink, J. L. (2012). Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PloS One* 7, e32253. doi: 10.1371/journal.pone.0032253

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

Rathan, N. D., Krishna, H., Ellur, R. K., Sehgal, D., Govindan, V., Ahlawat, A. K., et al. (2022). Genome-wide association study identifies loci and candidate genes for grain micronutrients and quality traits in wheat (*Triticum aestivum* L.). *Sci. Rep.* 12, 7037. doi: 10.1038/s41598-022-10618-w

Ray, D. K., Mueller, N. D., West, P. C., and Foley, J. A. (2013). Yield trends are insufficient to double global crop production by 2050. *PloS One* 8, e66428. doi: 10.1371/journal.pone.0066428

Ray, D. K., Ramankutty, N., Mueller, N. D., West, P. C., and Foley, J. A. (2012). Recent patterns of crop yield growth and stagnation. *Nat. Commun.* 3, 1293. doi: 10.1038/ncomms2296

Riaz, A., Athiyannan, N., Periyannan, S. K., Afanasenko, O., Mitrofanova, O. P., Platz, G. J., et al. (2018). Unlocking new alleles for leaf rust resistance in the Vavilov wheat collection. *Theor. Appl. Genet.* 131, 127–144. doi: 10.1007/s00122-017-2990-5

Ruppel, N. J., Logsdon, C. A., Whippo, C. W., Inoue, K., and Hangarter, R. P. (2011). A mutation in arabidopsis seedling plastid development1 affects plastid differentiation in embryo-derived tissues during seedling growth. *Plant Physiol.* 155 (1), 342–353. doi: 10.1104/pp.110.161414

Santana-Molina, C., Gutierrez, F., and Devos, D. P. (2021). Homology and modular evolution of CATCHR at the origin of the eukaryotic endomembrane system. *Genome Biol. Evol.* 13 (7), evab125. doi: 10.1093/gbe/evab125

Semnaninejad, H., Nourmohammadi, G., Rameeh, V., and Cherati, A. (2021). Correlation and path coefficient analyses of phenological traits, yield components and quality traits in wheat. *Rev. Bras. Engenharia Agricola e Ambiental* 25, 597–603. doi: 10.1590/1807-1929/agriambi.v25n9p597-603

Sheoran, S., Jaiswal, S., Kumar, D., Raghav, N., Sharma, R., Pawar, S., et al. (2019). Uncovering genomic regions associated with 36 agro-morphological traits in Indian spring wheat using GWAS. *Front. Plant Sci.* 10, 527. doi: 10.3389/fpls.2019.00527

Shuey, W. C. (1960). A wheat sizing technique for predicting flour milling yield. *Cereal Sci. Today* 5, 71.

Sourdille, P., Perretant, M. R., Charmet, G., Leroy, P., Gautier, M. F., Joudrier, P., et al. (1996). Linkage between RFLP markers and genes affecting kernel hardness in wheat. *Theor. Appl. Genet.* 93, 580–586. doi: 10.1007/BF00417951

Sukumaran, S., Dreisigacker, S., Lopes, M., Chavez, P., and Reynolds, M. P. (2015). Genome-wide association study for grain yield and related traits in an elite spring wheat population grown in temperate irrigated environments. *Theor. Appl. Genet.* 128, 353–363. doi: 10.1007/s00122-014-2435-3

Suprayogi, Y., Pozniak, C. J., Clarke, F. R., Clarke, J. M., Knox, R. E., and Singh, A. K. (2009). Identification and validation of quantitative trait loci for grain protein concentration in adapted Canadian durum wheat populations. *Theor. Appl. Genet.* 119, 437–448. doi: 10.1007/s00122-009-1050-1

Thorwarth, P., Ahlemeyer, J., Bochard, A. M., Krumnacker, K., Blümel, H., Laubach, E., et al. (2017). Genomic prediction ability for yield-related traits in German winter barley elite material. *Theor. Appl. Genet.* 130, 1669–1683. doi: 10.1007/s00122-017-2917-1

Tian, X., Engel, B. A., Qian, H., Hua, E., Sun, S., and Wang, Y. (2021). Will reaching the maximum achievable yield potential meet future global food demand? *J. Cleaner Prod.* 294, 126285. doi: 10.1016/j.jclepro.2021.126285

Tsenov, N., Gubatov, T., and Yanchev, I. (2020). Correlations between grain yield and related traits in winter wheat under multi-environmental traits. *Agric. Sci. Technol.* 12, 295–300. doi: 10.15547/issn1314-412X

Uauy, C., Distelfeld, A., Fahima, T., Blechl, A., and Dubcovsky, J. (2006). A NAC gene regulating senescence improves grain protein, zinc, and iron content in wheat. *Science* 314, 1298–1301. doi: 10.1126/science.1133649

van der Biezen, E. A., and Jones, J. D. (1998). The NB-ARC domain: a novel signalling motif shared by plant resistance gene products and regulators of cell death in animals. *Curr. Biol. CB* 8 (7), R226–R227. doi: 10.1016/s0960-9822(98)70145-9

van Ittersum, M. K., van Bussel, L. G. J., Wolf, J., Grassini, P., van Wart, J., Guilpart, N., et al. (2016). Can sub-Saharan Africa feed itself? *Proc. Natl. Acad. Sci. U. S. A.* 113, 14964–14969. doi: 10.1073/pnas.1610359113

Vishwakarma, M. K., Arun, B., Mishra, V. K., Yadav, P. S., Kumar, H., and Joshi, A. K. (2016). Marker-assisted improvement of grain protein content and grain weight in Indian bread wheat. *Euphytica* 208, 313–321. doi: 10.1007/s10681-015-1598-6

Vishwakarma, M. K., Mishra, V. K., Gupta, P. K., Yadav, P. S., Kumar, H., and Joshi, A. K. (2014). Introgression of the high grain protein gene Gpc-B1 in an elite wheat variety of Indo-Gangetic Plains through marker assisted backcross breeding. *Curr. Plant Biol.* 1, 60–67. doi: 10.1016/j.cpb.2014.09.003

Weis, C., Hückelhoven, R., and Eichmann, R. (2013). LIFEGUARD proteins support plant colonization by biotrophic powdery mildew fungi. *J. Exp. Bot.* 64 (12), 3855–3867. doi: 10.1093/jxb/ert217

Woese, C. R., Olsen, G. J., Ibba, M., and Söll, D. (2000). Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process. *Microbiol. Mol. Biol. Rev. MMBR* 64 (1), 202–236. doi: 10.1128/MMBR.64.1.202-236.2000

Woo, J. S., Imm, J. H., Min, C. K., Kim, K. J., Cha, S. S., and Oh, B. H. (2006). Structural and functional insights into the B30.2/SPRY domain. *EMBO J.* 25 (6), 1353–1363. doi: 10.1038/sj.emboj.7600994

Worland, A. J. (1996). The influence of flowering time genes on environmental adaptability in European wheats. *Euphytica* 89, 49–57. doi: 10.1007/BF00015718

Würschum, T., Leiser, W. L., Langer, S. M., Tucker, M. R., and Longin, C. F.H. (2018). Phenotypic and genetic analysis of spike and kernel characteristics in wheat reveals long-term genetic trends of grain yield components. *Theor. Appl. Genet.* 131, 2071–2084. doi: 10.1007/s00122-018-3133-3

Xu, X., Bai, G., Carver, B. F., and Shaner, G. E. (2005). A QTL for early heading in wheat cultivar Suwon 92. *Euphytica* 146, 233–237. doi: 10.1007/s10681-005-9017-z

Yang, Y., Chai, Y., Zhang, X., Lu, S., Zhao Z, Z., Wei, D., et al (2020). Multi-locus GWAS of quality traits in bread wheat: mining more candidate genes and possible regulatory network. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.01091

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

Zampieri, M., Ceglar, A., Dentener, F., and Toreti, A. (2017). Wheat yield loss attributable to heat waves, drought and water excess at the global, national and subnational scales. *Environ. Res. Lett.* 12, 064008. doi: 10.1088/1748-9326/aa723b

Zanke, C., Ling, J., Plieske, J., Kollers, S., Ebmeyer, E., Korzun, V., et al. (2014). Genetic architecture of main effect QTL for heading date in European winter wheat. *Front. Plant Sci.* 5. doi: 10.3389/fpls.2014.00217

Zhang, P., Guo., G., Wu., Q., Chen, Y., Xie, J., Lu, P., et al. (2020). Identification and fine mapping of spot blotch (Bipolaris sorokiniana) resistance gene Sb4 in wheat. *Theor. Appl. Genet.* 133, 2451–2459. doi: 10.1007/s00122-020-03610-3

Zhang, K., Tian, J., Zhao, L., Liu, B., and Chen, G. (2009). Detection of quantitative trait loci for heading date based on the doubled haploid progeny of two elite Chinese wheat cultivars. *Genetica* 135, 257–265. doi: 10.1007/s10709-008-9274-6

Zhang, C., Zheng, B., and He, Y. (2021). Improving grain yield via promotion of kernel weight in high yielding winter wheat genotypes. *Biology* 11, 42. doi: 10.3390/biology11010042

Zhu, C., Gore, M., Buckler, E. S., and Yu, J. (2008). Status and prospects of association mapping in plants. *Plant Genome* 1, 1. doi: 10.3835/plantgenome2008.02.0089

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 15 October 2024
DOI 10.3389/fpls.2024.1395938

Check for updates

OPEN ACCESS

EDITED BY
Reyazul Rouf Mir,
Sher-e-Kashmir University of Agricultural Sciences and Technology, India

REVIEWED BY
Vikas Venu Kumaran,
Indian Agricultural Research Institute (ICAR), India
Sandhya Tyagi,
United Arab Emirates University, United Arab Emirates
Mehboob-ur- Rahman,
National Institute for Biotechnology and Genetic Engineering, Pakistan

*CORRESPONDENCE
Rajendra Kumar
rajendrak64@yahoo.co.in
Aladdin Hamwieh
a.hamwieh@cgiar.org

RECEIVED 04 March 2024
ACCEPTED 09 September 2024
PUBLISHED 15 October 2024

CITATION
Chandana BS, Mahto RK, Singh RK, Bhandari A, Tandon G, Singh KK, Kushwah S, Lavanya GR, Iquebal MA, Jain N, Kudapa H, Upadhyaya HD, Hamwieh A and Kumar R (2024) Genome-wide association mapping identifies novel SNPs for root nodulation and agronomic traits in chickpea. *Front. Plant Sci.* 15:1395938. doi: 10.3389/fpls.2024.1395938

COPYRIGHT
© 2024 Chandana, Mahto, Singh, Bhandari, Tandon, Singh, Kushwah, Lavanya, Iquebal, Jain, Kudapa, Upadhyaya, Hamwieh and Kumar. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Genome-wide association mapping identifies novel SNPs for root nodulation and agronomic traits in chickpea

B. S. Chandana[1], Rohit Kumar Mahto[1,2], Rajesh Kumar Singh[1], Aditi Bhandari[3], Gitanjali Tandon[4], K. K. Singh[5], Sunita Kushwah[6], Gera Roopa Lavanya[7], Mir Asif Iquebal[4], Neelu Jain[1], Himabindu Kudapa[3], H. D. Upadhyaya[3,8], Aladdin Hamwieh[9]* and Rajendra Kumar[1]*

[1]Division of Genetics, Indian Agricultural Statistics Research Institute (ICAR)-Indian Agricultural Research Institute, New Delhi, India, [2]School of Biotechnology, Institute of Science, Banaras Hindu University (BHU), Varanasi, India, [3]International Crops Research Institute for the Semi-Arid Tropics, Patancheru, Telangana, India, [4]Division of Bioinformatics, ICAR-IASRI, New Delhi, India, [5]ICAR-Indian Agricultural Research Institute Regional Station, Samstipur, Bihar, India, [6]Krishi Vigyan Kendra (KVK), Vaishali, Dr. Rajendra Prasad Central Agriculture University-Pusa, Hajipur, Bihar, India, [7]Department of Genetics & Plant Breeding, Sam Higginbottom University of Agriculture, Technology and Sciences (SHUATS), Prayagraj, India, [8]Plant Genome Mapping Laboratory, University of Georgia, Athens, Athens, GA, United States, [9]International Center for Agriculture Research in the Dry Areas (ICARDA), Giza, Egypt

Introduction: The chickpea (Cicer arietinum L.) is well-known for having climate resilience and atmospheric nitrogen fixation ability. Global demand for nitrogenous fertilizer is predicted to increase by 1.4% annually, and the loss of billions of dollars in farm profit has drawn attention to the need for alternative sources of nitrogen. The ability of chickpea to obtain sufficient nitrogen via its symbiotic relationship with *Mesorhizobium ciceri* is of critical importance in determining the growth and production of chickpea.

Methods: To support findings on nodule formation in chickpea and to map the genomic regions for nodulation, an association panel consisting of 271 genotypes, selected from the global chickpea germplasm including four checks at four locations, was evaluated, and data were recorded for nodulation and 12 yield-related traits. A genome-wide association study (GWAS) was conducted using phenotypic data and genotypic data was extracted from whole-genome resequencing data of chickpea by creating a hap map file consisting of 602,344 single-nucleotide polymorphisms (SNPs) in the working set with best-fit models of association mapping.

Results and Discussion: The GWAS panel was found to be structured with sufficient diversity among the genotypes. Linkage disequilibrium (LD) analysis showed an LD decay value of 37.3 MB, indicating that SNPs within this distance behave as inheritance blocks. A total of 450 and 632 stringent marker–trait associations (MTAs) were identified from the BLINK and FarmCPU models, respectively, for all the traits under study. The 75 novel MTAs identified for nodulation traits were found to be stable. SNP annotations of associated markers were found to be related to various genes including a few auxins encoding as well

as nod factor transporter genes. The identified significant MTAs, candidate genes, and associated markers have the potential for use in marker-assisted selection for developing high-nodulation cultivars after validation in the breeding populations.

KEYWORDS

association mapping, chickpea, GWAS, nitrogen fixation, nodulation, PVE, pleiotropic

1 Introduction

Chickpea (*Cicer arietinum* L.) is a self-pollinated diploid crop with a chromosome number of 2n = 2x = 16, which is grown as an annual crop mainly during the winter season and is the third most important pulse crop globally with a cultivated area of 15.00 million hectares, production of 15.87 million tons, and average productivity of 1.06 t/ha (FAOSTAT, 2023). Chickpea along with other legumes can transform nitrogen from the atmosphere into ammonia through a symbiotic relationship with a rhizobium, *Mesorhizobium ciceri*. The ability of the chickpea to acquire adequate nitrogen through its symbiotic association with *M. ciceri* is essential for promoting growth and facilitating grain yield. Farmers exploit this mutually beneficial interaction with rhizobia to overcome nutrient deficiencies in soils, as these bacteria can supply as much as 97% of a plant's total nitrogen demand (Peoples and Craswell, 1992). In addition, these symbiotic relationships play a crucial role in replenishing substantial amounts of nitrogen in agricultural soils and thereby decreasing the reliance on expensive fertilizer treatments worldwide (Herridge et al., 2008). Gaining a greater understanding of the aspects that could enhance the advantages of this mutually beneficial relationship would be rewarding in the field of agriculture. Comprehending the relationship between genotype and nodulation in chickpea is crucial for optimizing the advantages of nitrogen fixation and minimizing the need for nitrogenous fertilizers. The whole process of symbiosis and nodulation is quite complex and tightly regulated and still has not been explored at the molecular level in chickpea. Nevertheless, a considerable number of genes associated with the process of nodulation at various stages have been identified in model legumes like *Medicago truncatula* and *Lotus japonicus*, employing a mix of forward and reverse genetics investigations (Roy et al., 2020). Several genes implicated in nodulation were initially discovered as nodulin genes that have elevated expression levels in nodules as compared to other plant tissues. Reverse genetics tests demonstrated that a significant number of these genes encoded proteins that played a role in nodulation (Combier et al., 2006). Precise improvement of complex quantitative traits like root nodulation traits needs the identification of related genomic regions rather than the identification of genes and quantitative trait locus (QTL) mapping, a robust technique that requires either bi-parental mapping populations, which is time-consuming (Edae et al., 2014), or genome-wide association study (GWAS) based on the linkage disequilibrium (LD) for the identification of genes/QTLs. Despite that chickpea is the most important food legume, nodulation studies in chickpea have been limited. Hence, a high-throughput, in-depth analysis of the chickpea root nodule is crucial for gaining deeper insights into the complexities of nodulation events. Identification of genotypes as resources for high nodulation and establishing an association between the nodulation traits and molecular markers can produce a higher yield per unit area. So far, the chickpea germplasm including the global core collection has not been fully utilized for the purpose. Thus, we conducted a systematic evaluation of conserved germplasm to facilitate the identification of high-nodulation genotypes with the objectives of phenotyping of nodulation and yield traits as resources, and we conducted a genome-wide association study to establish the association between the nodulation traits and molecular markers/genomic regions in chickpea.

2 Materials and methods

2.1 Plant material

A set of 2,094 diverse germplasms including a global core set of 1,950 genotypes and Indian Agricultural Research Institute (IARI) breeding materials (144) of chickpea was evaluated for the number of nodules and yield per se traits. The core germplasms collected from 28 different countries across the world were obtained from the gene bank at the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, Telangana, India. The plant materials were grown and evaluated for two consecutive crop seasons in 2018–2019 and 2019–2020 at IARI, New Delhi. Phenotypic data for nodulation and yield per se traits were recorded. Data were subjected to core hunter3 in R (De Brucellae et al., 2018) and descriptive statistics for the construction of four association panels (APs) focusing on nodulation (two APs), root (one AP), and plant architecture traits (one AP). The association panel under study consists of 271 diverse germplasm inclusive of BG 372, BG 3022, BG 547, and BG 1105 as four checks

(Supplementary Table 1). The experimental trials for the association panel were conducted at four environmental locations in 2020–2021, as follows: IARI, New Delhi, location 1 (28°38′24.0252″N latitude, 77°10′26.328″E longitude, and 228.6 m AMSL) having sandy clay loam soils; Sam Higginbottom University of Agriculture, Technology and Sciences (SHUATS), Naini, Prayagraj, location 2 (25°24′41.27″N latitude, 81°51′3.42″E longitude, and 98 m AMSL) with clay loam to sandy loam soil; Dr. Rajendra Prasad Central Agricultural University (RPCAU), Samastipur (KVK, Vaishali), location 3 (25°86′29.679″ latitude, 85°78′10.263″ longitude, and 52 m AMSL) with sandy loam soil; and IARI Regional Station, Pusa, Bihar, location 4 (25°54′56.16″ latitude, 85°40′24.9564″ longitude, and 52 m AMSL) with alluvial soils. The layout embodied an augmented randomized block design with four blocks and a spacing of 60 cm between rows and 10 cm between plants. Each block consisted of 72 lines including repeated check rows. The observations recorded on randomly selected five plants for each genotype for 12 traits included days to 50% flowering (DFF), plant height (PH) in cm, number of pods (NOP) per plant, number of seeds (NOS) per plant and yield (SY) per plant in grams, number of nodules (NON) per plant, nodule fresh weight (NFW) in grams, root fresh weight (RFW) in grams, root dry weight (RDW) in grams, stem fresh weight (SFW) in grams, and stem dry weight (SDW) in grams.

2.2 Phenotyping and data analysis

The nodule phenotyping pipeline includes mainly two parameters: counting the number of nodules per plant and taking nodule fresh weight as explained further. Phenotyping for the number of nodules was conducted 60 days after sowing; the optimum stage in legumes to fix maximum biological nitrogen was as reported earlier (Yuan et al., 2022) and followed the steps shown in Figure 1. Randomly selected five plants from each genotype were uprooted from the adhered soil mass using a hand hoe by digging 20 cm or even deeper into the soil (step 1). Particular care was taken not to disturb the root nodule system during sampling, and adhering soil was removed carefully (step 2). Root and shoot systems were separated (step 3). Roots with intact nodules were washed, and the number of nodules was counted (step 4). The intact cleaned roots were stored in butter paper bags to further obtain nodule and root fresh weight. The shoots were also kept in polythene packets. Phenotyping for root, shoot, and nodule fresh weight was conducted on the same day followed by their storage in the oven at 55°C for 1 week to obtain their dry weight. The explained procedure was followed for all the locations, the observations for all the traits were taken for five plants, and the mean of five plants per genotype was taken into consideration for analysis. Phenotypic data analysis including frequency distribution

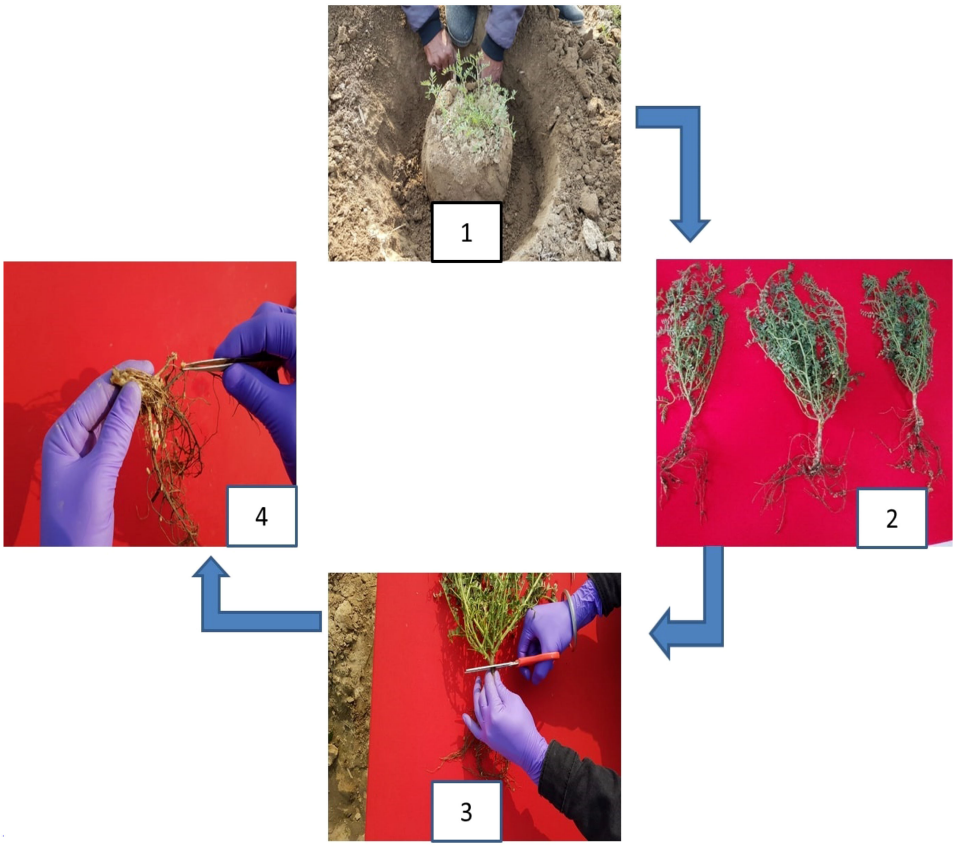


FIGURE 1
Phenotyping steps for number of nodules.

and correlation for all four locations was conducted using the R software (https://www.R-project.org/).

2.3 Genotyping of the association panel

Genotypic data for the association panel were successfully obtained from whole-genome resequencing of chickpea (Varshney et al., 2019). For single-nucleotide polymorphism (SNP), called clean reads were mapped on the reference genome of the chickpea genotype CDC Frontier (Varshney et al., 2013). To filter out low-quality variants, the loci with sequencing depth higher than 10,000 and lower than 400, mapping times higher than 1.5, or quality scores lower than 20 were used. The loci with estimated allele frequency not equal to 0 or 1 were determined as SNPs. The raw genotypic data extracted from the database contained 1,198,121 SNPs distributed on eight pseudomolecules. The filtering for missing data (≤20%) and minor allele frequency (MAF) ≥2% was performed using vcf tools (Vogt et al., 2021); an additional filter for the rate of heterozygosity (Ho) ≤ 0.5%, MAF ≥ 5%, and Ho ≤ 5% led to a working set of 602,344 SNPs (referred as 602K), which were used for genome-wide association mapping analysis.

2.4 Association analysis

The generated genotyping data were integrated with phenotypic data of multi-location observations recorded for the traits under study. Phenotypic data were used for the calculation of best linear unbiased predictions (BLUPs). Individual BLUPs across the environments were estimated using the ACBD-R software (Rodríguez et al., 2017) with the following model:

$$Yij = \mu + Genj + Envi + Envi \times Genj + Bloc(Envi) + eij$$

where Genj and Checkj correspond to the effects of the identifier of checks, the un-replicated genotypes, and checks that are repeated in each block (Blocki); Envi is the effect of ith environment, μ is the mean, and e is the error component (as described in ACBD-R User Manual; Rodríguez et al., 2017). The population structure was assessed using a neighbor-joining phylogenetic tree (constructed through the TASSEL software and visualized through the ITOL software) and principal component analysis (PCA). PCA was performed using a function dedicated to assessing the genetic relatedness among accessions and generating the principal components (PCs) from the genotypic data. The first three principal components were considered as covariates in GAPIT using the high-performance computing R tool. The r^2 values for SNP markers were computed and then filtered focusing on pairs within each chromosome, and a linkage disequilibrium heat map was created to identify significant LD block and its size, which falls diagonally in the heat map at a p-value of 0.001. A whole genome was generated and sorted for individual chromosomes by utilizing TASSEL version 5. Subsequently, these files were used to generate LD decay curves for all eight chromosomes individually and for the entire genome. To estimate the sizes of LD blocks, the r^2 values were plotted against the distance in base pairs (bp) while setting a threshold at $r^2 = 0.2$.

GWAS was performed using the general linear model (GLM), mixed linear model (MLM), multi-locus mixed model (MLMM), FarmCPU model, and BLINK model using the R/GAPIT 3.0 package. Further, in this study, the Bonferroni correction threshold value of −log10 > 7.0 (p-value) was used as the cutoff. The SNPs with the above values were declared as significant marker–trait associations (MTAs). The Manhattan and Q-Q plots were generated through qqman version 0.1.8 (Turner, 2014). The percent phenotypic variance (PV) explained by all significant detected SNPs was generated from all used models and calculated as the squared correlation between the phenotype and genotype of the SNP. Stable MTAs obtained more than twice across the location were found. Pleotropic SNPs having an association with more than one trait were also identified.

2.5 Identification of associated SNPs and candidate genes

The genes involving significant SNP markers were aligned against the National Center for Biotechnology Information (NCBI) non-redundant (nr) protein database using BLASTX to obtain functional annotations (https://blast.ncbi.nlm.nih.gov). The stable and pleiotropic SNPs were subjected to a basic local alignment search tool (BLAST) search using the sequence information of the markers. A BLAST search was carried out using a data web service. Putative candidate transcripts (with transcript IDs) within and 20-kb flanking region of SNPs were identified in the NCBI chickpea database, and the function of the gene was determined using the UniProt database (https://www.uniprot.org/).

3 Results

3.1 Distribution and correlation among the nodulation traits

Phenotypic data collected under all the environments for the association panel were statically analyzed, and the results are presented further. The mean and distributions for 12 phenotypic variables in 271 accessions chosen from the chickpea reference set are presented in Figure 2. The traits under study exhibited normal and near-normal to skewed distributions. The mean of the DFF (53.42), Days to maturity (DTM) (143.84), PH (42.74), NOP (84.27), NOS (74.32), yield per plant (8.24), NON (8.83), NFW (336.61), RFW (2.16), RDW (0.46), SFW (6.69), and SDW (4.23) were recorded. The results of correlation coefficients revealed that nodule fresh weight and nodule dry weight were positively and significantly correlated with yield $plant^{-1}$ at genotypic and phenotypic levels (Figure 3).

The results of the correlation coefficients evidenced that the genotypic correlations for most of the traits were slightly higher than their corresponding phenotypic correlations, which would be

10.3389/fpls.2024.1395938

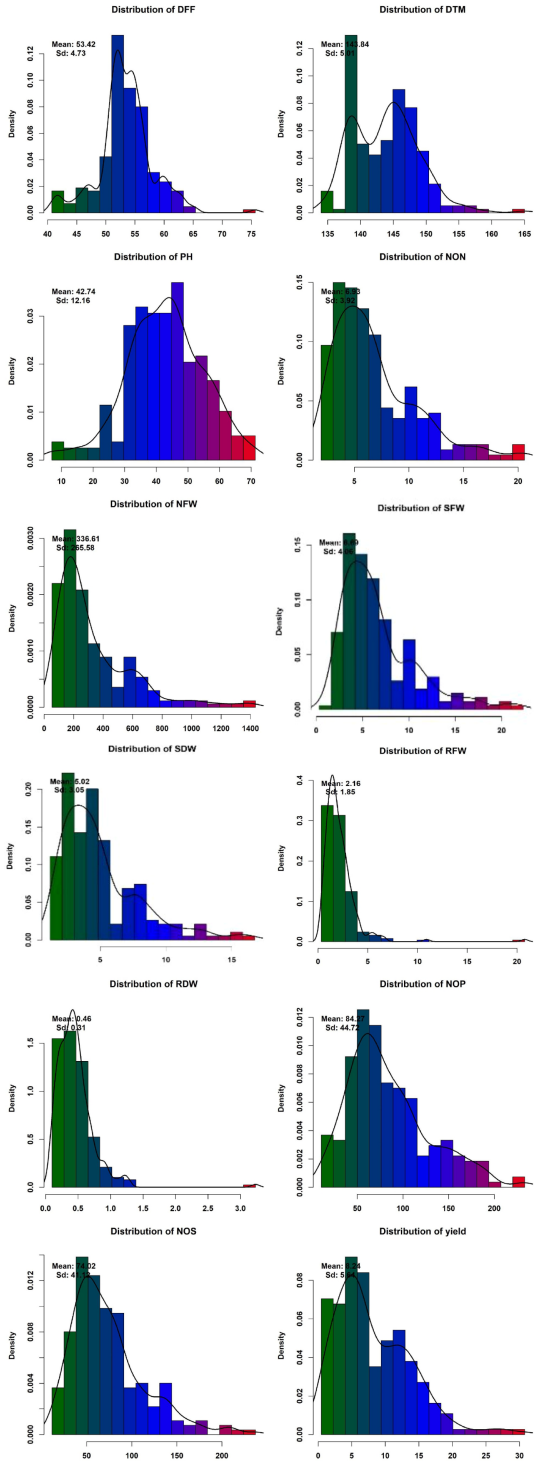


FIGURE 2
Phenotypic variation for traits assayed within the chickpea reference set.

beneficial in the selection of traits because they exclude the environmental influence. It also revealed significant and positive correlation values for seed yield with the number of pods per plant, seeds per plant, shoot fresh weight, days to flowering, and days to maturity. However, negative correlation values for seed yield with shoot dry weight and nodule fresh weight were observed. Significant and positive correlations were observed for the trait SY with NPB, NSB, and NPP, indicating that the seed yield may be enhanced through an increase in Number of primary branches (NPB), Number of secondary branches (NSB), and Number of pods per plant (NPP).

3.2 Assessment of population structure and linkage disequilibrium block

In order to assess the number of subpopulations, a phylogenetic tree utilizing phenotypic and marker data through the neighbor-joining method was constructed (Figure 4A). The phylogenetic tree revealed the presence of three subpopulations/subclusters, which were further confirmed by the generated PCA scree plot (Figure 4B). Subcluster 3 was the largest one, containing 202 inclusive of all checks, followed by subcluster 2 containing 55 and subcluster 1 containing 14 genotypes. Subclusters 1 and 2 remained confined to ICC series germplasm lines.

The LD across the genome was estimated using 603,100 SNPs from the working set through the TASSEL software for the whole genome. The average LD across the genome was 635.9 kb. The distribution of SNPs across the chromosome was on eight different pseudomolecules in chickpea, as presented in Figure 5A. The number of SNPs available on each pseudomolecule and the number of SNPs used for conducting marker–trait association are represented in Figure 5B. The genomic regions represented in dark red on the chromosome were found to have a high density of SNPs, and the genomic regions represented in green had low SNP density. Ca 4 had the highest number of SNPs, and Ca 8 contained the least number of SNPs.

3.3 Genome-wide association study for identification of MTAs for nodulation and agronomic traits

A GWAS analysis was carried out to identify SNPs associated with investigated variables in chickpeas, including nodulation, morphological, and yield traits. The GWAS identified total SNPs through BLINK (643), FarmCPU (720), and MLM (439) models; the number of significant identified SNPs for each trait under different models in four different conditions are listed in Table 1; additional SNPs identified for each trait, along with their chromosome, position and Phenotypic variance explained (PVE) identified through the Blink model, FarmCPU model, and MLM are given in Supplementary Tables 2-11. As the current study focused on nodulation traits number of nodules and nodule fresh weight, their stable SNPs are presented in Table 2, and Manhattan and Q-Q plots are represented in Supplementary Figures 1 and 2. We mainly considered SNPs that had a p-value threshold of $-\log10$ p-value ≥ 6 and a false discovery rate (FDR) below 0.1. The GWAS for the trait NON found eight SNPs that are stably expressible at locations 1 and 2; one SNP that is stably expressible at locations 1, 2, and 3; and one SNP that is stably expressible at locations 1, 2, and 4 (Table 2). Among these identified MTAs, SNP 2_825902 had 27.33% of PVE. Marker–trait association of nodule fresh weight resulted in the

10.3389/fpls.2024.1395938

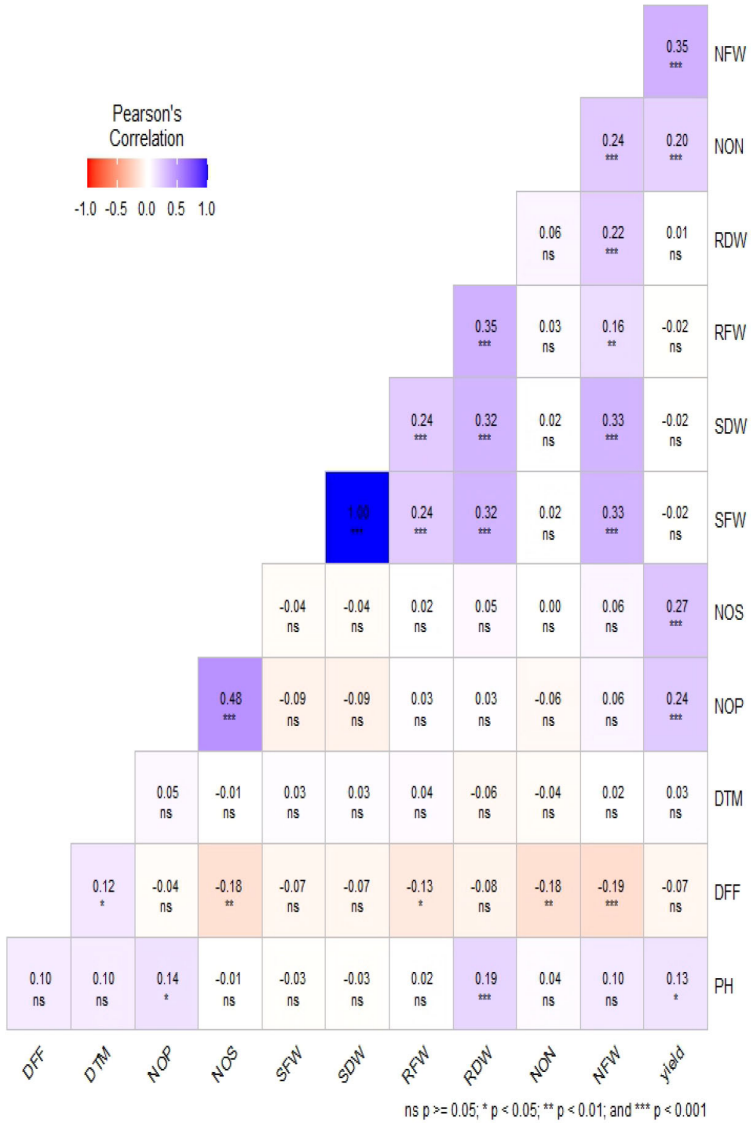
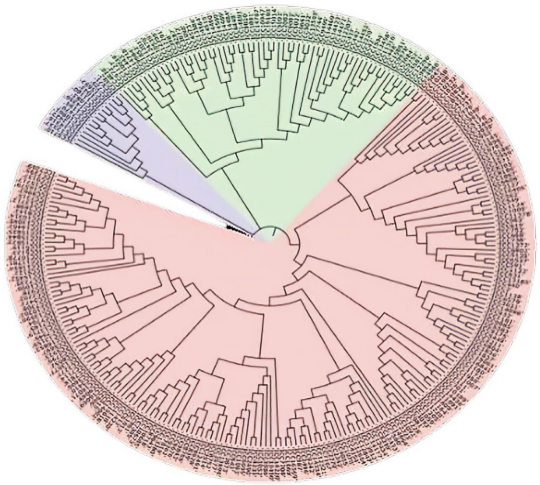


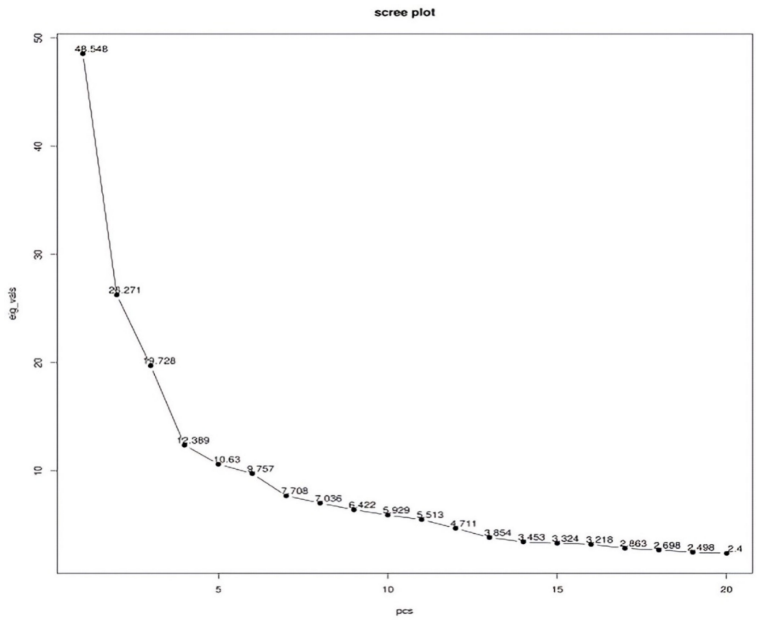
FIGURE 3
Estimation of Pearson's correlation coefficients for the chickpea association panel.

identification of SNP markers as 20 in the FarmCPU model; among the significant identified markers, seven SNP markers were found above the threshold of –log10 p-value. SNPs Ca5pos20514758.1 and Ca7pos21461047.1 presented on chromosome numbers 5 and 7 had 42.28% and 9.31% of PVE, respectively, through FarmCPU.

The list of the stable SNPs and their location along with their trait are in Table 3. Stable SNPs are SNPs that were identified in more than one environment. Most of the SNPs were stably expressive and common for location 3 *vs.* 4 and location 1 *vs.* 3. For plant height, we found 57, 72, and 29 SNPs in the BLINK model, FarmCPU model, and MLMM, respectively. Among the identified markers, we report 12 stable SNPs. SNP Ca6pos1492432.1 (FarmCPU for location 1) found on chromosome number 6 had 25.98% of PVE, and SNP Ca2pos6782498.2 (BLINK, location 3) present on chromosome number 2 had 27.24% of PVE (Supplementary Tables 2-10). SNP Ca8pos11994085.1 present on chromosome 8 had 14.56% of PVE (the SNPs along with PVE% are given in Supplementary Tables 2-11). We identified 14 stable SNPs each for DFF, DTM, and PH. The stable SNPs for DFF were identified only from the MLM; however, for DTM and plant height, stable SNPs were common across the models (MLM, BLINK, and FarmCPU), and for all three traits, we observed stable SNPs at locations 3 and 4. Among these MTAs, the stable SNP for the trait DFF was Ca6pos47821883.1, present on chromosome number 6 and had 48.43% of PVE (MLM and GLM). SNP Ca1pos4393831.1 present on chromosome number 1 had 8.82% of PVE. SNP Ca5pos30670011.1 present on chromosome number 5 had 9.15% of Phenotypic variance explained (PVE). The trait DTM contained SNPs Ca1pos9168435.1 and Ca1pos11296743.1 present on chromosome number 1, which had 42.37% and 33.81% of PVE, respectively; SNPs Ca5pos13855141.1 and Ca5pos30670011.1 present on chromosome 5 number had 8.05% and 9.15% of PVE, respectively. For the traits SFW and NOS, we identified 20 stable SNPs. SNPs Ca1pos33781183.1 and



(a)



(b)

FIGURE 4
(A) Phylogenetic tree. **(B)** PCA scree plot. PCA, principal component analysis.

Ca1pos35026875.1 identified for trait SFW had 8.91% and 3.7% of PVE, respectively. SNPs Ca5pos30670011.1, Ca3pos19977000.1, and Ca4pos27656718.1 identified for NOS had 9.15%, 7.04%, and 3.74% of PVE, respectively.

3.5 SNP markers associated with two or more different traits

An important initial step in the process of revealing pleiotropic loci associated with complex phenotypes is to examine SNPs that have already been independently associated with one or more different traits using the statistically stringent GWAS framework. The GWAS results were fully examined in order to identify markers that were common between traits (Table 4). A total of 43 SNPs were found common, with notable p-values and statistically significant FDRs that reduce the chances of false associations and increase the chances of true association discovered in this study. There were four markers found common for RFW *vs.* RDW and 16 markers for SFW *vs.* SDW, and 13 markers were present on chromosome 3 between genomic regions 37990876 and 38114728. Nineteen markers were common for NPP *vs.* Number of seeds per plant

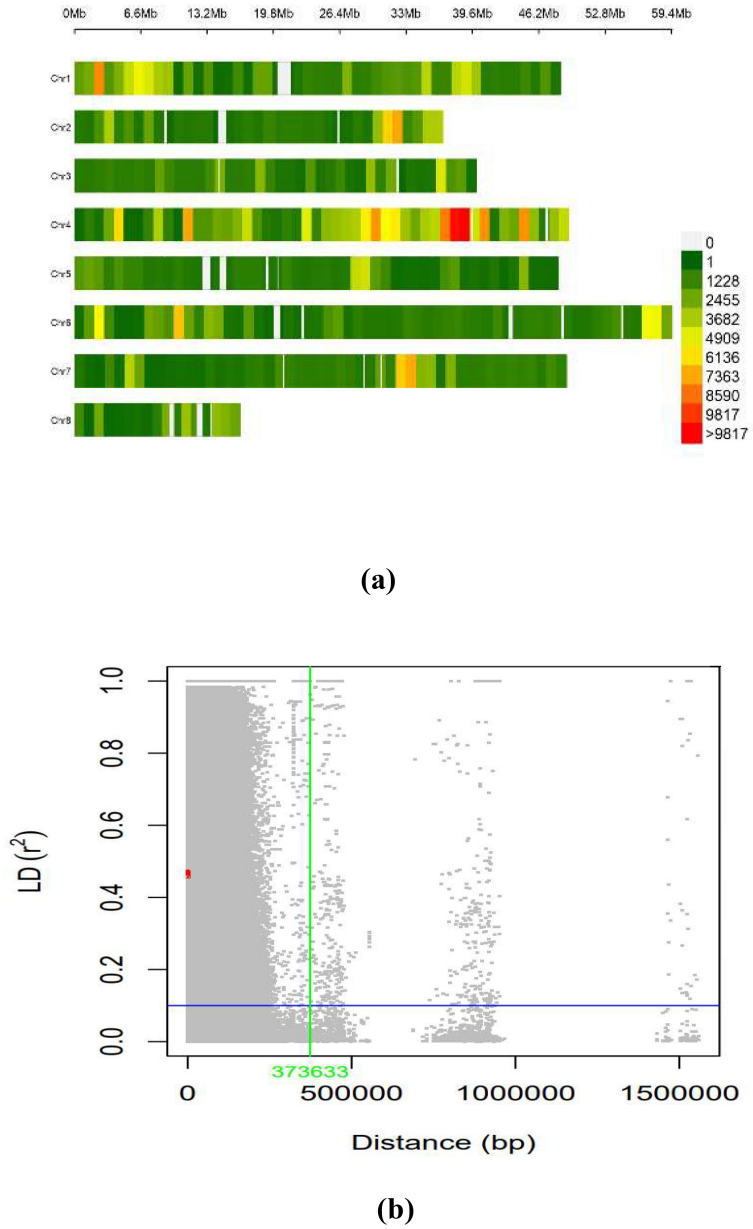


FIGURE 5
(A) SNP density plot indicating distribution of filtered SNPs across the chromosomes. **(B)** Linkage disequilibrium measured r^2 plotting *vs.* physical distance between pairs of markers (Plink 1.9). SNP, single-nucleotide polymorphism.

(NSP) and were distributed across all eight chromosomes of the chickpea.

3.6 Identification of putative candidate genes for associated SNPs

A BLAST search of significant SNPs identified from the current association panel was aligned against the CDC Frontier reference genome of chickpea, revealing the location of SNPs near the gene-rich region of the genome but present in intergenic regions. Irrespective of being in intron regions, the majority of the SNPs were near one or the other transcripts coding for some proteins or transcription factors. Many SNPs were located near the genes coding for general proteins well known for their metabolic function in growth and development like membrane proteins, DNA/RNA recognition/binding protein, ABC transporter, and protein kinase. SNP markers like 1_10074058 and 1_28905467 were linked to the genes governing the traits that are indirectly involved in root nodulation like auxin-responsive protein, environmental stress, and plant–microbe interaction candidate genes near the location of identified MTAs; their transcript IDs, proteins produced from them, and role of those proteins in plants based on previous reports are given in Tables 5A and Table 5B.

10.3389/fpls.2024.1395938

TABLE 1 List of the significant SNP markers identified using different models.

Trait	Blink		FarmCPU		MLM	
	SNP	-Log 10 (p)	SNP	-Log 10 (p)	SNP	-Log 10 (p)
DFF	88	4 to 8	76	4 to 8	57	4 to 8
DTM	76	4 to 8	80	4 to 8	60	4 to 8
PH	57	4 to 6	72	4 to 6	29	4 to 6
NON	12	7 to 14	2	7 to 14	3	7 to 14
NFW	15	7 to 14	15	7 to 14	4	7 to 14
RFW	60	4 to 8	80	4 to 8	75	4 to 8
RDW	58	4 to 8	78	4 to 8	72	4 to 8
SFW	35	4 to 8	37	4 to 8	20	4 to 8
SDW	22	4 to 8	45	4 to 8	19	4 to 8
NPP	70	4 to 8	80	4 to 8	35	4 to 8
NSP	80	4 to 8	80	4 to 8	30	4 to 8
SY	70	4 to 8	75	4 to 8	35	4 to 8
Total	643		720		439	

The list of all the identified markers along with chromosome location, PVE, and position of the markers are given in Supplementary Tables 2-11, and Manhattan plots are given in Supplementary Figures 1-10.

DFF, days to 50% flowering; PH, plant height (cm); PB, primary branches; SB, secondary branches; NOP, number of pods per plant; NOS, number of seeds per plant; Yield, yield per plant (g); NON, number of nodules per plant; NFW, nodule fresh weight (g); RFW, root fresh weight (g); RDW, root dry weight (g); SFW, stem fresh weight (g); SDW, stem dry weight (g); SNP, single-nucleotide polymorphism.

4 Discussions

4.1 Genome-wide association analysis as a competitive tool for identification of genomic regions controlling chickpea nodulation traits

The chickpea crop has a narrow genetic base (Kumar et al., 2017; Singh et al., 2022). Due to the use of a limited number of accessions/donor parents in breeding programs, the sensitivity of chickpea productivity toward biotic and abiotic stress is increased (Muehlbauer and Sarker, 2017). Developing high-nodulation genotypes will help in the augmentation of chickpea potentiality. In this investigation, on the basis of frequency distribution, it may be concluded that a wide range of variability existed for NON, NFW, RFW, RDW, and other yield-contributing traits. Further, correlation coefficients of our association panel revealed that nodule fresh weight and nodule dry weight were positively and significantly correlated with yield plant^{-1} at genotypic and phenotypic levels. These results are in accordance with those of several studies (Chandana et al., 2023; Elias, 2009; Mohamed and Hassan, 2015). This indicated that the development of effective and promising nodules helps in increasing yield. This was probably due to the uptake of atmospheric nitrogen through the process of biological nitrogen fixation. To accelerate the breeding programs, conventional breeding efforts need to be augmented using genomics-assisted breeding (Varshney et al., 2007, 2013). In our study, we carried out resequencing by whole-genome resequencing, which resulted in the identification of

TABLE 2 Significant marker–trait associations at Bonferroni correction p-value for traits under study at different locations/environments.

Trait	SNP ID	Chromosome	Position (bp)	p-Value	MAF	Location
NON	Ca2pos2169123.1	2	2169123	4.22E−08	0.00369	1
	Ca2pos28892511.1	2	2.90E+07	6.33E−15	0.09963	1 and 2
	Ca2pos28892511.1	2	2.90E+07	5.96E−09	0.09963	1 and 2
	Ca3pos36960027.1	3	3.70E+07	2.64E−08	0.04982	1 and 2
	Ca4pos43172963.1	4	4.30E+07	5.61E−12	0.02768	1 and 2
	Ca4pos43172963.1	4	4.30E+07	1.63E−09	0.02768	1, 2, and 4
	Ca5pos2026241.1	5	2026241	5.86E−15	0.49631	1 and 2
	Ca5pos21727560.1	5	2.20E+07	3.26E−14	0.49815	1, 2, and 3
	Ca5pos21727560.1	5	2.20E+07	9.49E−18	0.49815	1 and 2
	Ca5pos38432936.1	5	3.80E+07	4.22E−08	0.49815	1 and 2
	Ca7pos32196399.1	7	3.20E+07	3.26E−10	0.04428	1 and 2
	Ca7pos47087116.1	7	4.70E+07	7.67E−08	0.19202	2
NFW	Ca7pos33694339.1	7	3.40E+07	6.25E−14	0.15926	1 and 3
FarmCPU	Ca3pos8244189.1	3	8244189	8.44E−11	0.13889	1 and 3
	Ca8pos11345025.1	8	1.10E+07	1.51E−08	0.09815	1 and 3
	Ca6pos56503783.1	6	5.70E+07	4.85E−08	0.0037	1 and 3

(Continued)

10.3389/fpls.2024.1395938

TABLE 2 Continued

Trait	SNP ID	Chromosome	Position (bp)	p-Value	MAF	Location
	Ca2pos34684339.1	2	3.50E+07	5.23E−08	0.02222	1 and 3
	Ca4pos40834901.1	4	4.10E+07	5.52E−08	0.21482	1 and 3
	Ca6pos43064226.1	6	4.30E+07	1.34E−07	0.20185	1 and 3
	Ca4pos4851043.1	4	4851043	3.65E−07	0.04259	1 and 3
	Ca3pos13147329.1	3	1.30E+07	5.01E−07	0.40185	1 and 3
	Ca8pos12988095.1	8	1.30E+07	1.55E−06	0.4963	1 and 4
	Ca8pos11435804.1	8	1.10E+07	2.94E−06	0.04815	1 and 4
	Ca7pos21461047.1	7	2.10E+07	3.85E−06	0.39815	1, 2, and 4
	Ca4pos2034216.1	4	2034216	4.22E−06	0.49259	1, 2, and 4
	Ca5pos20514758	5	3.30E+07	7.71E−06	0.33704	1, 2, and 4
	Ca4pos29978493.1	4	3.00E+07	9.61E−06	0.09074	1 and 4
	Ca3pos8244095.1	3	8244095	9.64E−06	0.05926	1 and 3

MAF, minor allele frequency; SNP, single-nucleotide polymorphism.

TABLE 3 Stable SNPs associated with more than one environment.

DFF					
SNP	Linkage group	Position	p-Value	Model	Location
Ca1pos24610010.1	1	24610010	7.50E−05	MLM	3 and 4
Ca2pos1198412.1	2	1198412	1.35E−05	MLM	3 and 4
Ca2pos8664225.1	2	8664225	8.93E−05	MLM	3 and 4
Ca3pos12432325.1	3	12432325	8.59E−05	MLM	3 and 4
Ca3pos30297117.1	3	30297117	7.05E−05	MLM	3 and 4
Ca6pos16026122.1	6	16026122	8.45E−05	MLM	3 and 4
Ca6pos41345425.1	6	41345425	9.11E−05	MLM	3 and 4
Ca6pos41972856.1	6	41972856	7.30E−05	MLM	3 and 4
Ca6pos41972896.1	6	41972896	8.50E−06	MLM	3 and 4
Ca6pos48048350.1	6	48048350	8.38E−05	MLM	3 and 4
Ca7pos13030083.1	7	13030083	5.33E−05	MLM	3 and 4
Ca7pos19732025.1	7	19732025	3.66−05	MLM	3 and 4
Ca7pos38107103.1	7	38107103	4.55E−05	MLM	3 and 4
Ca7pos43994019.1	7	43994019	9.52E−05	MLM	3 and 4
DTM					
SNP	Chromosome	Position	p-Value	Model	Location
Ca1pos29737724.1	1	29737724	1.55E−05	MLM, BLINK, and FarmCPU.	3 and 4
Ca1pos29737735.1	1	29737735	1.83E−05		3 and 4
Ca3pos433121.1	3	433121	0.00010207		3 and 4
Ca3pos10956124.1	3	10956124	0.00011024		3 and 4
Ca3pos13147329.1	3	13147329	3.76E−05		3 and 4

(Continued)

10.3389/fpls.2024.1395938

TABLE 3 Continued

DTM					
SNP	Chromosome	Position	p-Value	Model	Location
Ca5pos29910669.1	5	29910669	7.19E−05		3 and 4
Ca5pos29910672.1	5	29910672	1.21E−05		3 and 4
Ca6pos32512792.1	6	32512792	6.62E−05		3 and 4
Ca6pos34164845.1	6	34164845	6.28E−05		3 and 4
Ca6pos49267149.1	6	49267149	6.13E−05		3 and 4
Ca6pos56003699.1	6	56003699	9.10E−05		3 and 4
Ca6pos58002384.1	6	58002384	2.65E−07		3 and 4
Ca7pos30208448.1	7	30208448	9.94E−05		3 and 4
Ca7pos37819116.1	7	37819116	2.06E−05		3 and 4
Plant height					
SNP	Chromosome	Position	p-Value	Model	Location
Ca1pos27773061.1	1	27773061	8.25E−05	MLM and FarmCPU	3 and 4
Ca2pos20166203.1	2	20166203	9.11E−05	MLM and FarmCPU	3 and 4
Ca3pos2766368.1	3	2766368	6.92E−05	MLM	3 and 4
Ca4pos8732683.1	4	8732683	7.41E−05	MLM	3 and 4
Ca4pos8818268.1	4	8818268	2.51E−05	MLM and FarmCPU	3 and 4
Ca4pos9195726.1	4	9195726	2.20E−05	MLM and FarmCPU	3 and 4
Ca4pos35809754.1	4	35809754	6.44E−05	MLM	3 and 4
Ca5pos1941712.1	5	1941712	6.74E−05	MLM	3 and 4
Ca5pos5415849.1	5	5415849	3.72E−05	MLM	3 and 4
Ca6pos42565470.1	6	42565470	5.02E−05	MLM	3 and 4
Ca6pos47675915.1	6	47675915	1.36E−05	MLM	3 and 4
Ca6pos47675923.1	6	47675923	3.74E−05	MLM and FarmCPU	3 and 4
Ca6pos52464995.1	6	52464995	5.38E−05	MLM and FarmCPU	3 and 4
Ca7pos30981168.1	7	30981168	9.33E−05	MLM	3 and 4
Ca7pos40493709.1	7	40493709	4.56E−05	MLM and FarmCPU	3 and 4
SFW					
Ca1pos3076296.1	1	3076296	2.81E−05	MLM	3 and 4
Ca1pos3105289.1	1	3105289	4.73E−05	MLM	3 and 4
Ca1pos37122468.1	1	37122468	2.91E−05	MLM	3 and 4
Ca1pos39647846.1	1	39647846	1.54E−05	MLM	3 and 4
Ca2pos28823139.1	2	28823139	4.73E−05	MLM	3 and 4
Ca3pos19707707.1	3	19707707	3.93E−05	MLM	3 and 4
Ca3pos25805689.1	3	25805689	6.09E−05	MLM	3 and 4
Ca3pos25805690.1	3	25805690	6.09E−05	MLM	3 and 4
Ca4pos32772485.1	4	32772485	2.04E−05	MLM	3 and 4
Ca5pos33032029.1	5	33032029	3.79E−05	MLM	3 and 4
Ca5pos41523197.1	5	41523197	6.74E−05	MLM	3 and 4

(Continued)

TABLE 3 Continued

SFW					
Ca5pos42012947.1	5	42012947	3.52E−05	MLM	3 and 4
Ca5pos42074441.1	5	42074441	1.16E−05	MLM	3 and 4
Ca5pos42272729.1	5	42272729	2.09E−05	MLM	3 and 4
Ca6pos14512673.1	6	14512673	2.23E−05	MLM	3 and 4
Ca6pos27009414.1	6	27009414	5.12E−05	BLINK and FarmCPU	3 and 4
Ca7pos15917660.1	7	15917660	2.39E−05	BLINK and FarmCPU	3 and 4
Ca7pos30666345.1	7	30666345	2.04E−05	BLINK and FarmCPU	3 and 4
Ca8pos12985289.1	8	12985289	4.32E−05	BLINK and FarmCPU	3 and 4
NOS					
Ca1pos8737037.1	1	8737037	1.46E−05	BLINK and FarmCPU	1 and 3
Ca1pos8820515.1	1	8820515	4.41E−05	BLINK and FarmCPU	1 and 3
Ca1pos14062422.1	1	14062422	3.50E−06	BLINK and FarmCPU	1 and 3
Ca1pos14070458.1	1	14070458	5.92E−06	BLINK and FarmCPU	1 and 3
Ca1pos46718466.1	1	46718466	3.87E−05	BLINK and FarmCPU	1 and 3
Ca2pos11422406.1	2	11422406	2.13E−05	BLINK, FarmCPU, and MLM	1 and 3
Ca2pos11422414.1	2	11422414	3.55E−05	BLINK, FarmCPU, and MLM	1 and 3
Ca3pos3042321.1	3	3042321	3.86E−05	BLINK and FarmCPU	1 and 3
Ca3pos39844991.1	3	39844991	9.40E−06	BLINK and FarmCPU	1 and 3
Ca3pos39885397.1	3	39885397	1.26E−05	BLINK and FarmCPU	1 and 3
Ca4pos19410981.1	4	19410981	8.00E−06	BLINK and FarmCPU	1 and 3
Ca5pos3834038.1	5	3834038	7.49E−06	BLINK and FarmCPU	1 and 3
Ca5pos46549190.1	5	46549190	8.46E−06	BLINK, FarmCPU, and MLM	1 and 3
Ca6pos24093514.1	6	24093514	2.61E−05	BLINK and FarmCPU	1 and 3
Ca6pos28350598.1	6	28350598	3.44E−05	BLINK and FarmCPU	1 and 3
Ca6pos53644742.1	6	53644742	3.20E−05	BLINK and FarmCPU	1 and 3
Ca7pos27650319.1	7	27650319	4.45E−05	BLINK, FarmCPU, and MLM	1 and 3
Ca7pos47164247.1	7	47164247	1.35E−05	BLINK and FarmCPU	1 and 3
Ca8pos13311157.1	8	13311157	1.15E−05	BLINK and FarmCPU	1 and 3

SNP, single-nucleotide polymorphism; DFF, days to 50% flowering; SFW, stem fresh weight (g); NOS, number of seeds per plant.

1,198,121 SNPs. The sequencing cost was curtailed as an advance in the next-generation sequencing technology (Varshney et al., 2009), and the availability of chickpea draft genome (Jain et al., 2013; Varshney et al., 2013; Gupta et al., 2017) led to the identification of several millions of SNPs and other molecular markers. Further integrating whole-genome resequencing information with precise phenotyping provides detailed information about all genetic variants and enables the discovery of novel variations.

4.2 Diversity and linkage disequilibrium analysis for identifying MTAs close to the trait of interest

In GWAS analysis, population structure may be a confounding factor that must be addressed to avoid false associations. Population structure and uneven relatedness among the individuals of the population act as confounding factors and lead to spurious identification of MTAs (Zhang et al., 2010). Admixture of PCA

10.3389/fpls.2024.1395938

TABLE 4 List of the common SNP markers found for more than one trait (pleiotropic SNPs).

	SNP	Chromosome	Position	p-Value	Location	Traits
1	Ca1pos34151789.1	1	34151789	6.68E–05	3 and 4	RFW and RDW filtering criteria
2	Ca6pos38012786.1	6	38012786	4.43E–05	3 and 4	RFW and RDW
3	Ca7pos19898936.1	7	19898936	6.01E–05	3 and 4	RFW and RDW
4	Ca7pos22479713.1	7	22479713	3.20E–06	3 and 4	RFW and RDW
5	Ca1pos2169207.1	1	2474918	1.23E–06	3 and 4	SFW and SDW
6	Ca1pos2474918.1	1	40820287	1.69E–06	3 and 4	SFW and SDW
7	Ca1pos40820287.1	1	37940720	3.46E–06	3 and 4	SFW and SDW
8	Ca3pos37940720.1	3	37990876	9.19E–06	3 and 4	SFW and SDW
9	Ca3pos37990876.1	3	38008999	5.25E–06	3 and 4	SFW and SDW
10	Ca3pos38057183.1	3	38057193	1.85E–06	3 and 4	SFW and SDW
11	Ca3pos38057193.1	3	38057228	7.16E–06	3 and 4	SFW and SDW
12	Ca3pos38057228.1	3	38063432	8.30E–06	3 and 4	SFW and SDW
13	Ca3pos38063432.1	3	38063441	4.21E–06	3 and 4	SFW and SDW
14	Ca3pos38063441.1	3	38068021	7.06E–06	3 and 4	SFW and SDW
15	Ca3pos38068021.1	3	38080272	4.01E–06	3 and 4	SFW and SDW
16	Ca3pos38080272.1	3	38100180	2.29E–06	3 and 4	SFW and SDW
17	Ca3pos38100180.1	3	38100212	5.60E–06	3 and 4	SFW and SDW
18	Ca3pos38100212.1	3	38106616	7.44E–06	1, 3, and 4	SFW and SDW
19	Ca3pos38106616.1	3	38114728	7.45E–07	1, 3, and 4	SFW and SDW
20	Ca3pos38114728.1	3	6667809	8.90E–06	1, 3, and 4	SFW and SDW
21	Ca7pos6667809.1	7	46119789	4.42E–06	1, 3, and 4	SFW and SDW
22	Ca1pos8737036.1	1	8737036	1.46E–05	1 and 3	NSP and NPP
23	Ca1pos8737037.1	1	8737037	1.46E–05	1 and 3	NSP and NPP
24	Ca1pos8820515.1	1	8820515	4.41E–05	1 and 3	NSP and NPP
25	Ca1pos14062422.1	1	14062422	3.50E–06	1 and 3	NSP and NPP
26	Ca1pos14070458.1	1	14070458	5.92E–06	1 and 3	NSP and NPP
27	Ca1pos46718466.1	1	46718466	3.87E–05	1 and 3	NSP and NPP
28	Ca2pos11422406.1	2	11422406	2.13E–05	1, 2, and 4	NSP and NPP
29	Ca2pos11422414.1	2	11422414	3.55E–05	1, 2, and 4	NSP and NPP
30	Ca3pos3042321.1	3	3042321	3.86E–05	1, 2, and 4	NSP and NPP
31	Ca3pos39844991.1	3	39844991	9.40E–06	1, 2, and 4	NSP and NPP
32	Ca3pos39885397.1	3	39885397	1.26E–05	1, 2, and 4	NSP and NPP
33	Ca4pos19410981.1	4	19410981	8.00E–06	1, 2, and 4	NSP and NPP
34	Ca5pos3834038.1	5	3834038	7.49E–06	1, 2, and 4	NSP and NPP
35	Ca5pos46549190.1	5	46549190	8.46E–06	1, 2, and 4	NSP and NPP
36	Ca6pos24093514.1	6	24093514	2.61E–05	3 and 4	NSP and NPP
37	Ca6pos28350598.1	6	28350598	3.44E–05	3 and 4	NSP and NPP
38	Ca6pos53644742.1	6	53644742	3.20E–05	3 and 4	NSP and NPP

(Continued)

TABLE 4 Continued

	SNP	Chromosome	Position	p-Value	Location	Traits
39	Ca7pos27650319.1	7	27650319	4.45E–05	3 and 4	NSP and NPP
40	Ca7pos47164247.1	7	47164247	1.35E–05	3 and 4	NSP and NPP
41	Ca8pos13311157.1	8	13311157	1.15E–05	3 and 4	NSP and NPP
	SNP	Chromosome	Position	p-Value	Location	Traits
1	Ca1pos34151789.1	1	34151789	6.68E–05	3 and 4	RFW and RDW
2	Ca6pos38012786.1	6	38012786	4.43E–05	3 and 4	RFW and RDW
3	Ca7pos19898936.1	7	19898936	6.01E–05	3 and 4	RFW and RDW
4	Ca7pos22479713.1	7	22479713	3.20E–06	3 and 4	RFW and RDW
5	Ca1pos2169207.1	1	2474918	1.23E–06	3 and 4	SFW and SDW
6	Ca1pos2474918.1	1	40820287	1.69E–06	3 and 4	SFW and SDW
7	Ca1pos40820287.1	1	37940720	3.46E–06	3 and 4	SFW and SDW
8	Ca3pos37940720.1	3	37990876	9.19E–06	3 and 4	SFW and SDW
9	Ca3pos37990876.1	3	38008999	5.25E–06	3 and 4	SFW and SDW
10	Ca3pos38057183.1	3	38057193	1.85E–06	3 and 4	SFW and SDW
11	Ca3pos38057193.1	3	38057228	7.16E–06	3 and 4	SFW and SDW
12	Ca3pos38057228.1	3	38063432	8.30E–06	3 and 4	SFW and SDW
13	Ca3pos38063432.1	3	38063441	4.21E–06	3 and 4	SFW and SDW
14	Ca3pos38063441.1	3	38068021	7.06E–06	3 and 4	SFW and SDW
15	Ca3pos38068021.1	3	38080272	4.01E–06	3 and 4	SFW and SDW
16	Ca3pos38080272.1	3	38100180	2.29E–06	3 and 4	SFW and SDW
17	Ca3pos38100180.1	3	38100212	5.60E–06	3 and 4	SFW and SDW
18	Ca3pos38100212.1	3	38106616	7.44E–06	1, 3, and 4	SFW and SDW
19	Ca3pos38106616.1	3	38114728	7.45E–07	1, 3, and 4	SFW and SDW
20	Ca3pos38114728.1	3	6667809	8.90E–06	1, 3, and 4	SFW and SDW
21	Ca7pos6667809.1	7	46119789	4.42E–06	1, 3, and 4	SFW and SDW
22	Ca1pos8737036.1	1	8737036	1.46E–05	1 and 3	NSP and NPP
23	Ca1pos8737037.1	1	8737037	1.46E–05	1 and 3	NSP and NPP
24	Ca1pos8820515.1	1	8820515	4.41E–05	1 and 3	NSP and NPP
25	Ca1pos14062422.1	1	14062422	3.50E–06	1 and 3	NSP and NPP
26	Ca1pos14070458.1	1	14070458	5.92E–06	1 and 3	NSP and NPP
27	Ca1pos46718466.1	1	46718466	3.87E–05	1 and 3	NSP and NPP
28	Ca2pos11422406.1	2	11422406	2.13E–05	1, 2 and 4	NSP and NPP
29	Ca2pos11422414.1	2	11422414	3.55E–05	1, 2 and 4	NSP and NPP
30	Ca3pos3042321.1	3	3042321	3.86E–05	1, 2 and 4	NSP and NPP
31	Ca3pos39844991.1	3	39844991	9.40E–06	1, 2 and 4	NSP and NPP
32	Ca3pos39885397.1	3	39885397	1.26E–05	1, 2 and 4	NSP and NPP
33	Ca4pos19410981.1	4	19410981	8.00E–06	1, 2 and 4	NSP and NPP
34	Ca5pos3834038.1	5	3834038	7.49E–06	1, 2 and 4	NSP and NPP
35	Ca5pos46549190.1	5	46549190	8.46E–06	1, 2 and 4	NSP and NPP

(Continued)

TABLE 4 Continued

	SNP	Chromosome	Position	p-Value	Location	Traits
	SNP	Chromosome	Position	p-Value	Location	Traits
36	Ca6pos24093514.1	6	24093514	2.61E–05	3 and 4	NSP and NPP
37	Ca6pos28350598.1	6	28350598	3.44E–05	3 and 4	NSP and NPP
38	Ca6pos53644742.1	6	53644742	3.20E–05	3 and 4	NSP and NPP
39	Ca7pos27650319.1	7	27650319	4.45E–05	3 and 4	NSP and NPP
40	Ca7pos47164247.1	7	47164247	1.35E–05	3 and 4	NSP and NPP
41	Ca8pos13311157.1	8	13311157	1.15E–05	3 and 4	NSP and NPP

SNP, single-nucleotide polymorphism; RFW, root fresh weight (g); RDW, root dry weight (g); SFW, stem fresh weight (g); SDW, stem dry weight (g).

and phylogenetic tree using the neighbor-joining method are popular approaches for accurately inferring population structure derived from the genome-wide association panels using high-density SNPs (Abraham and Inouye, 2014). The use of genotypes from various origins of the world in the study may be the reason for the three clear-cut subpopulations in the GWAS panel (Upadhyaya et al., 2001). In neighbor-joining dendrograms, the different clustering patterns are clearly visible, and similar results were reported by Thudi et al. (2021). The use of a diverse panel of genotypes can provide more valuable inference compared to bi-parental populations (Vos-Fels et al., 2017) by taking advantage of maximum allelic diversity and historic recombination events; the present study materials having diverse lines with their distant parentage ensure the required diversity for the association study. The genetic diversity of chickpea germplasm could provide important information for selecting effective parental breeding strategies as well as a better understanding of natural variations in phenotypic traits and their genetic background (Raina et al., 2019). The genome-wide LD decay estimated for the present investigation is on par with the previous association studies of chickpea for complex yield traits (Upadhyaya et al., 2016; Kujur et al., 2015). An LD value of 373 kb illustrates the string association of SNP markers with the trait of interest. The SNP density analysis revealed that SNPs are distributed across all the chromosomes, and this helps to identify all possible causal variants for traits of interest. This aligns with a previous study on Desi and Kabuli chickpea genotypes and gives the importance of geographic origin and adaptive environments in genotype clustering (Basu et al., 2018; Roorkiwal et al., 2022). The major challenge for GWAS is to control the false positives primarily caused by population structure and family relatedness (Kaler et al., 2020). To avoid such bias, a stringent selection procedure of the Bonferroni correction was applied; a total of 643, 720, and 439 SNPs were identified in the association panel through the BLINK model, FarmCPU model, and MLM, respectively. Identification of such a huge number of SNPs can provide deeper insights into the genomic regions associated with traits of interest. In this study, our main target was nodulation traits such as NON, NFW, and other associated traits like shoot and root because the association between shoot- and root-related traits indirectly contributes to the nodulation. Hence, the identification of genomic regions for root and shoot regions will also help in the simultaneous improvement of the legume crop.

4.3 MTAs controlling nodulation and yield contributing traits

In the marker and nodulation trait association, we obtained 12 SNPs in the MLMM for NON-above FDR or –log10 p-value of 7, which were highly significant markers for the number of nodules and SNPs Ca4pos29278891.1 and Ca4pos34659769.1, identified through FarmCPU for location 1 and had phenotypic effects of 3.27% and 4.08%, respectively. As of now, there are no reported large-scale studies on the identification of SNPs for the number of nodules in chickpea; these SNPs will serve as markers for molecular studies and marker-assisted breeding in chickpea. For NON, we found highly significant MTAs on chromosome numbers 2 (three SNPs), 3 (one SNP), 4 (four SNPs), and 7 (two SNPs) through the MLMM. These findings indicate that an improved understanding of these markers and genomic regions of chickpea is necessary to improve the benefits of rhizobial symbiosis in chickpea root nodulation. The trait NFW, an important nodulation trait, was used to measure the nodulation potential of legumes, especially in the case of chickpeas, as they produce indeterminate nodules and also have a direct association with chickpea growth attributes and biological nitrogen fixation (Hazra et al., 2021). Istanbuli et al. (2024) reported for NFW in their study, the maximum number of SNPs on chromosome number 4 in genomic regions was between 52- and 58-Mb regions. However, in our study, we report 75 novel markers for the nodulation traits. SNP Ca5pos20514758.1 reported for NFW and presented on chromosome number 5 had 42.28% of PVE. The SNP was also based on gene annotation in the NCBI of the Cicer genome and was found to be an associated gene that encodes for important proteins like auxin-induced proteins, which are involved in nod factor binding export proteins in *Medicago* (Gully et al., 2018). Plant height in chickpeas is influenced by 200 SNPs; additionally, it is consistent with earlier studies showing multiple loci controlling plant height in different crops. The multi-locus control of plant height is influenced by both genetic and environmental factors, although specific details are not fully

TABLE 5A Putative candidate genes identified for NON at the 10-kb region of linked SNPs along with their molecular functions.

SNP	Start	End	Sequence description	Biological function
1_10074058	10071267	10077948	Auxin-responsive protein IAA26-like	Regulation of transcription, auxin-activated signaling pathway, main hormone for root nodule initiation
1_19310421	19307421	19312421	SNF1-related protein kinase regulatory subunit beta-3	Regulation of protein kinase activity, cellular response to nitrogen levels, response to sucrose
2_825900	805900	845900	Transmembrane protein, putative	
2_12931949	12913554	12916946	3392	Emp24/gp25L/p24 family protein
	12920791	12923570	2779	Hypothetical protein VIGAN_08277400
4_2035604	2031504	2036704	Adenosine kinase 2	Purine ribonucleoside salvage, AMP biosynthetic process, phosphorylation
6_16019517	16042593	16042985	Thermospermine synthase ACAULIS5-like	
6_16019517	16062478	16063311	RuBisCO-associated protein-like	Carbohydrate metabolic process
6_16019517	16075075	16075869	Probable UDP-arabinopyranose mutase 5	Plant-type cell wall organization or biogenesis

understood (Yang et al., 2021). As provided in Supplementary Tables 2-11, traits such as DTM, DTF, RFW, RDW, SFW, SDW, NPP, NSP, and yield were controlled by more than at least 130 SNP markers. Among them for the trait DTF, the significant SNPs were Ca4pos99711.1 with 2.8% of PVE and Ca5pos30670011.1 with 9.15% of PVE identified for location 1. SNP Ca7pos16225558, which is identified in both the FarmCPU model MLM, had 5.94% of PVE. SNP Ca6pos47821883.1 identified for location 3 had 48.43% of PVE.

Through GWAS, we were able to identify highly significant markers for NFW through different models as mentioned in the Results section. Therefore, these markers serve indirectly in the improvement of nodulation in chickpea, which supports the establishment and survival of plants in adverse conditions, thereby promoting plant productivity (Gupta et al., 2015); also, our analyses provide new insights into the identities of markers and phenotypic influences on identified markers by providing causal variants for responsible markers that explained phenotypic variance. The MTAs identified in more than one environment were grouped as stable/promising ones, while MTAs with greater than 15% phenotypic variation were grouped as major MTAs. The major MTAs found in our study identify a large number of stable SNPs. Thus, the GWAS results in the identification of stable markers for all the traits such as four SNPs for RFW, seven SNPs for SFW, and 17 stable SNPs for SDW, which may be regarded as nodulation-related traits, as these traits play a role in root development and establishment. The precise phenotyping and high accuracy may lead to a greater number of stable expressions along the different locations/conditions presumed to be the real association of these markers for governing the phenotypic expression of the studied traits. Limited stable SNPs for root traits through GWAS have also been reported earlier (Thudi et al., 2014). Pleiotropy occurs when a single genetic variant is associated with multiple phenotypic outcomes. Hence, it is common to find markers to be associated with more than one trait, i.e., pleiotropic influence. Significant pleiotropic loci were detected for nodulation and yield. Thus, we can improve both

TABLE 5B Putative candidate genes identified for NFW at the 10-kb region of linked SNPs along with their molecular functions.

SNP	Protein	Function	References
1_29852482	FT-interacting protein 7-like	Regulators of plant immunity	Ashraf et al., 2018
	FT-interacting protein 7-like	Plant immunity	
1_28905467	BZIP transcription factor TGA10-like isoform X2	Environmental stress	Li et al., 2020
	FT-interacting protein 7-like	Vernalization response	Gretsova et al., 2023
6_3345569	60S ribosomal protein L34	Alternate spicing	Denise et al., 2019
	PsbP-like protein 1, chloroplast isoform X2	Assembly of the PSII SCs is required for adaptation to changing light intensity	Che et al., 2020
8_11345025	50S ribosomal protein L23, chloroplast	Support translation under stress	
8_11435804	F-box protein At2g39490-like	Disease resistance	Zhou et al., 2013
8_12988095	Transcription initiation factor TFIID subunit 6	Plant–microbe interactions	Ning et al., 2021

NFW, nodule fresh weight (g); SNPs, single-nucleotide polymorphisms.

traits and the productivity of the crop simultaneously and indirectly. Similarly, we also found different stable MTAs showing pleiotropic effects between different yield and nodulation traits due to their interdependence on each other. These traits are bound to share genes in common, and the results are supported by the presence of a significant positive correlation between nodulation and yield traits in our study. The PVE% explained by SNPs refers to the proportion of the total variability observed in a specific trait or phenotype that can be attributed to genetic variations in SNPs. PVE % is a measure of the contribution of genetic factors to a particular trait. The SNPs with high PVE can be considered as major and significant for that particular trait. Two SNPs were found to be stable at locations 1, 3, and 4; seven SNPs were found to be stable at locations 1 and 3 for the trait NON. SNP 2_825902, which was found on chromosome number 2, had 27.33% of the PVE. Also, for the other traits, we identified many good SNPs whose number and percentage of PVE are already explained in the Results section. Stable SNPs with high PVE can be used as fixed effects in the genomic selection pipeline.

4.4 Gene enrichment analysis for nodulation traits

In the current investigation, as our main focus was on nodulation traits, we tried to provide genes that are located within the identified MTAs for nodulation traits, and the investigation revealed MTAs within a prominent genomic region housing candidate genes responsible for governing diverse functions in plant growth, developmental processes, biotic and abiotic pathways, stress tolerance, and intricate nodulation pathways. Particularly noteworthy is the stably identified MTA 1_10074058 associated with the trait of interest NON. This locus encodes an auxin-responsive protein IAA26-like, a pivotal hormone instrumental in root initiation and initiation of root nodules. Furthermore, its involvement in transcriptional regulation and auxin-activated signaling pathways underscores its significance (Luo et al., 2018). Similarly, the SNP 1_19310421 associated with the NON trait is situated in a genomic region where the gene codes for an SNF1-related protein kinase regulatory subunit beta-3. This subunit is implicated in the regulation of protein kinase activity, cellular responses to nitrogen levels, and the intricate response to sucrose signaling (Jamsheer et al., 2021). These findings collectively contribute to our understanding of the genetic underpinnings of the observed traits and the multifaceted processes governing plant–nutrient interactions and developmental responses.

5 Conclusion

Biological nitrogen fixation (BNF) stands for a sustainable and globally applicable avenue for supplying nitrogen to agricultural systems. An effective strategy to augment BNF involves the breeding and utilization of legume varieties possessing enhanced BNF capacity. Notably, our results demonstrated the effectiveness of phenotyping with increased row-to-plant spacing in elucidating traits related to nodule formation. Leveraging association studies, we successfully identified noteworthy and stable MTAs linked to the traits of interest. A total of more than 800 MTAS have been reported for root, shoot, yield, and other morphological traits. These MTAs can be compared in future studies to identify the most probable location of causal variation for their respective traits. Subsequent *in silico* analysis unveiled that a substantial proportion of these MTAs were situated within intergenic regions, with the potential to modulate genes associated with the focal traits. The seven novel SNPs, namely, Ca2pos2169123.1, Ca2pos28892511.1, Ca3pos36960027.1, Ca4pos43172963.1, Ca5pos2026241.1, Ca5pos21727560.1, and Ca7pos47087116.1, identified for the number of nodules were highly significant and found common in different models of GWAS were considered as SNPs probably controlling the genomic regions for the trait NON, as these MTAs are present near the genes that aid in biological nitrogen fixation pathway. SNP Ca5pos20514758.1 identified for NFW in locations 1 and 3 was found with a PVE of 42.28% and could be used to construct the cleaved amplified polymorphic sequence (CAPS) or Kompetitive allele specific PCR (KASP) markers that can serve as a PCR-based marker for identifying polymorphism for nodulation traits in chickpea and other legumes. The stable SNPs characterized by a high proportion of PVE will be integrated as fixed effects within the genomic selection pipeline, accentuating their potential impact on future breeding efforts. The identified candidate genes could be exploited in marker-assisted breeding, genomic selection, and genetic engineering to improve the nodulation efficiency in legumes and other crop species.

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

CS: Writing – review & editing, Software, Writing – original draft, Investigation, Data curation. RM: Writing – original draft, Investigation, Data curation. RS: Writing – original draft, Investigation, Data curation. AB: Writing – review & editing, Software, Formal Analysis, Data curation. GT: Writing – review & editing, Software, Formal Analysis. KS: Data curation, Writing – review & editing, Investigation. SK: Data curation, Writing – review & editing, Investigation. GL: Data curation, Writing – review & editing, Investigation. MI: Writing – review & editing, Software, Formal Analysis. NJ: Writing – review & editing, Software, Formal Analysis. HK: Writing – review & editing, Software, Formal Analysis. HU: Data curation, Writing – review & editing, Resources. AH: Writing – review & editing, Visualization, Methodology. RK: Writing – review & editing, Writing – original draft, Visualization, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Acknowledgments

The corresponding authors sincerely acknowledge the DST-SERB for the financial support by granting the project CRG/2019/006273 entitled utilization of germplasm for identification of genes/QTLs for nodulation through association mapping in chickpea. All authors also sincerely acknowledge the ICAR-IARI, New Delhi; ICAR-IARI Regional Station, Pusa, Bihar; KVK, Vaishali (RPCAU, Samstipur), Bihar; and Department of Genetics and Plant Breeding, SHUATS, Naini, Prayagraj, Uttar Pradesh, for providing facilities to carry out this research work.

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 Reviewer VK declared a shared affiliation with the authors CS, RM, RS, GT, MI, NJ and RK to the handling editor at the time of the review.

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2024.1395938/full#supplementary-material

References

Abraham, G., and Inouye, M. (2014). Fast principal component analysis of large-scale genome-wide data. *PloS One.9* 4, e93766. doi: 10.1371/journal.pone.0093766

Ashraf, N., Basu, S., and Narula, K. (2018). Integrative network analyses of wilt transcriptome in chickpea reveal genotype dependent regulatory hubs in immunity and susceptibility. *Scientific reports* 8, 6528. doi: 10.1038/s41598-018-19919-5

Basu, U., Srivastava, R., Bajaj, D., Thakro, V., Daware, A., Malik, N., et al. (2018). Genome-wide generation and genotyping of informative SNPs to scan molecularsignatures for seed yield in chickpea. *Sci. Rep.* 8, 13240. doi: 10.1038/s41598-018-29926-1

Chandana, B. S., Mahto, R. K., Singh, R. K., Singh, K. K., Kushwah, S., Lavanya, G. R., et al. (2023). Unclenching the potentials of global core germplasm for rootnodulation traits for increased biological nitrogen fixation andproductivity in chickpea (*Cicer arietinum* L.). *Indian J. Genet. Plant Breed.* 83, 526–534. doi: 10.31742/ISGPB.83.4.9

Che, Y., Kusama, S., Matsui, S., Suorsa, M., Nakano, T., Aro, E. M., et al. (2020). Arabidopsis PsbP-Like Protein Facilitates the Assembly of the Photosystem II Super complexes and Optimizes Plant Fitness under Fluctuating Light. *Plant & cell physiology* 61(6), 1168–1180. doi: 10.1093/pcp/pcaa045

Combier, J. P., Frugier, F., De Billy, F., Boualem, A., El-Yahyaoui, F., and Moreau, S. (2006). MtHAP2–1 is a key transcriptional regulator of a symbiotic nodule development regulated by microRNA169 in *Medicago truncatula*. *Genes Devel* 20, 3084–3088. doi: 10.1101/gad.402806

De Brucellae, H., Davenport, G. F., and Fack, V. (2018). Core Hunter 3: flexible core subset selection. *BMC bioinf.* 19, 203. doi: 10.1186/s12859-018-2209-z

Denise, R., Abby, S. S., and Rocha E., P. C. (2019). Diversification of the type IV filament superfamily into machines for adhesion, protein secretion, DNA uptake and motility. *PloS Biol.* 17 (7), e3000390. doi: 10.1371/journal.pbio.3000390

Edae, E. A., Byrne, P. F., Haley, S. D., Lopes, M. S., and Reynolds, M. P. (2014). Genome-wide association mapping of yield and yield components of spring wheat under contrasting moisture regimes. *Theor. Appl. Genet.* 127, 791–807. doi: 10.1007/s00122-013-2257-8

Elias, N. (2009). Optimizing chickpea nodulation for nitrogen fixation and yield in north eastern, Australia. New South wales, Australia: University of Western Sydney. doi: 10.1038/nature00842

FAOSTAT. (2023). (Rome). Available online at: https://www.fao.org/faostat/es/#data/QCL (Accessed 23 December 2023).

Gretsova, M., Surkova, S., Kanapin, A., Samsonova, A., Logacheva, M., Shcherbakov, A., et al. (2023). Transcriptomic analysis of flowering time genes in cultivated chickpea and wild cicer. *Int. J. Mol. Sci.* 24 (3), 2692. doi: 10.3390/ijms24032692

Gully, D., Czernic, P., Cruveiller, S., Mahé, F., Longin, C., Vallenet, D., et al. (2018). Transcriptome profiles of nod factor-independent symbiosis in the tropical legume *aeschynomeneevenia*. *Sci. Rep.* 8, 10934. doi: 10.1038/s41598-018-29301-0

Gupta, S., Kumar, T., Verma, S., Bharadwaj, C., and Bhatia, S. (2015). Development of gene-based markers for use in construction of the chickpea (*Cicer arietinum* L.) genetic linkage map and identification of QTLs associated with seed weight and plant height. *Mol. Bio. Rep.* 42, 1571–1580. doi: 10.1007/s11033-015-3925-3

Gupta, S., Nawaz, K., Parween, S., Roy, R., Sahu, K., Pole, A. K., et al. (2017). Draft genome sequence of *Cicer reticulatum*L.the wild progenitor of chickpea provides a resource for agronomic trait improvement. *DNA. Res.* 24, 1–10. doi: 10.1093/dnares/dsw042

Hazra, K. K., Nath, C. P., and Singh, S. S. (2021). Categorization of chickpea nodules and their relation with plant growth. *Natl. Acad. Sci. Lett.* 44, 91–95. doi: 10.1007/s40009-020-00972-6

Herridge, D. F., Peoples, M. B., and Boddey, R. M. (2008). Global inputs of biological nitrogen fixation in agricultural systems. *Plant Sci. J.* 311, 1–18. doi: 10.1007/s11104-008-9668-3

Istanbuli, T., Nassar, A. E., Abd El-Maksoud, M. M., Tawkaz, S., Alsamman, A. M., and Hamwieh, A. (2024). Genome-wide association study reveals SNP markers controlling drought tolerance and related agronomic traits in chickpea across multiple environments. *Front. Plant Sci.* 15. doi: 10.3389/fpls.2024.1260690

Jain, M., Misra, G., Patel, R. K., Priya, P., Jhanwar, S., Khan, A. W., et al. (2013). A draft genome sequence of the pulse crop chickpea (*Cicer arietinum* L.). *Plant.J* 74, 715–729. doi: 10.1111/tpj.2013.74.issue-5

Jamsheer, K. M., Kumar, M., and Srivastava, V. (2021). SNF1-related protein kinase 1: the many-faced signaling hub regulating developmental plasticity in plants. *J. Exp. Bot.* 72, 6042–6065. doi: 10.1093/jxb/erab079

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, 1794. doi: 10.3389/fpls.2019.01794

Kujur, A., Bajaj, D., Upadhyaya, H. D., Das, S., Ranjan, R., Shree, T., et al. (2015). A genome-wide SNP scan accelerates trait-regulatory genomic loci identification in chickpea. *Sci. Rep.* 5, 11166. doi: 10.1038/srep11166

Kumar, R., Yadav, R., Soi, S., Srinivasan,, Yadav, S. S., Mishra, J. P., et al. (2017). Morpho-molecular characterization of landraces / wild genotypes of cicer for biotic / abiotic stresses. *Legume Res.* 40, 974–984. doi: 10.18805/lr.v0iOF.9100

Li, W., Zhao, D., Dong, J., Kong, X., Zhang, Q., Li, T., et al. (2020). Atrtp5 negatively regulates plant resistance to phytophthora pathogens by modulating the biosynthesis of endogenous jasmonic acid and salicylic acid. *Mol. Plant Pathol.* 21, 95–108. doi: 10.1111/mpp.12883

Luo, J., Zhou, J. J., and Zhang, J. Z. (2018). Aux/IAA gene family in plants: molecular structure, regulation, and function. *Int. J. Mol. Sci.* 19, 259. doi: 10.3390/ijms19010259

Mohamed, A. A., and Hassan, M. A. (2015). Evaluation of two chickpea (*cicer arietinum*L.) cultivars in response to three rhizobium strains at river nile state, Sudan. *Merit Res. J. Agric. Sci. Soil Sci.* 3 (5), 062–069. Available online at: http://meritresearchjournals.org/asss/index.htm

Muehlbauer, F. J., and Sarker, A. (2017). "Economic importance of chickpea: Production, value, and world trade," in The chickpea genome *compendium of plant genomes* (Springer, Cham). doi: 10.1007/978-3-319-66117-9_2

Ning, H., Yang, S., Fan, B., Zhu, C., and Chen, Z. (2021). Expansion and functional diversification of TFIIB-like factors in plants. *Int. J. Mol. Sci.* 22 (3), 1078. doi: 10.3390/ijms22031078

Peoples, M. B., and Craswell, E. T. (1992). Biological nitrogen fixation: investments, expectations and actual contributions to agriculture. *Plant Soil*. 141, 13–39. doi: 10.1007/BF00011308

Raina, A., Khan, S., Wani, M. R., Laskar, R. A., and Mushtaq, W. (2019). "Chickpea (*Cicer arietinum* L.) cytogenetics, genetic diversity and breeding," in *Advances in Plant Breeding Strategies: Legumes*, vol. 7. (Springer: Springer Nature Switzerland), 53–112.

Rodríguez, F., Pérez, P., de los Campos, G., Alvarado, G., Pacheco, A., Crossa, J., et al. (2017). *BGLR (Bayesian Generalized Linear Regression) Version 2.0* Vol. 06 (CYMMYT, Mexico: International Maize and Wheat Improvement Centre), 20. Available online at: https://hdl.handle.net/11529/10848.

Roorkiwal, M., Bhandari, A., Barmukh, R., Bajaj, P., Valluri, V. K., Chitikineni, A., et al. (2022). Genome-wide association mapping of nutritional traits for designing superior chickpea varieties.Front. *Plant Sci.* 13. doi: 10.3389/fpls.2022.843911

Roy, S., Liu, W., Nandety, R. S., Crook, A., Mysore, K. S., and Pislariu, C. I. (2020). Celebrating 20 years of genetic discoveries in legume noduation and symbiotic nitrogen fixation. *Plant Cell* 3, 15–41. doi: 10.1105/tpc.19.00279

Singh, R. K., Singh, C. A., Chandana, B. S., Mahto, R. K., Patial, R., Gupta, A., et al. (2022). Exploring chickpea germplasm diversity for broadening the genetic base utilizing genomic resources. *Front. Genet.* 13. doi: 10.3389/fgene.2022.905771

Thudi, M., Chen, Y., Pang, J., Kalavikatte, D., Bajaj, P., Roorkiwal, M., et al. (2021). Novel genes and genetic loci associated with root morphological traits, phosphorus-acquisition efficiency and phosphorus-use efficiency in chickpea. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.636973

Thudi, M., Upadhyaya, H. D., Rathore, A., Gaur, P. M., Krishnamurthy, L., Roorkiwal, M., et al. (2014). Genetic dissection of drought and heat tolerance in chick pea through genome-wide and candidate gene-based association mapping approaches. *PloS One* 9, 96758. doi: 10.1371/journal.pone.0096758

Turner, S. (2014). qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *Biorxiv* 14, 005165. doi: 10.1101/005165

Upadhyaya, H. D., Bajaj, D., Narnoliya, L., Das, S., Kumar, V., Gowda, C. L., et al. (2016). Genome-wide scans for delineation of candidate genes regulating seed-protein content in chickpea. *Front. Plant Sci.* 7, 302. doi: 10.3389/fpls.2016.00302

Upadhyaya, H. D., Bramel, P. J., and Singh, S. (2001). Development of a chickpea core subset using geographic distribution and quantitative traits. *Crop Sci.* 41 (1), 206–210.

Varshney, R. K., Hoisington, D. A., Upadhyaya, H. D., Gaur, P. M., Nigam, S. N., Saxena, K., et al. (2007). "Molecular genetics and breeding of grain legume crops for the semi-arid tropics," in *Genomics-assisted crop improvement*, vol. II . Eds. R. K. Varshney and R. Tuberosa 2, 207–242.

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

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

Vogt, F., Shirsekar, G., and Weigel, D. (2021). vcf2gwas-python API for comprehensive GWAS analysis using GEMMA 38, 1–4. doi: 10.1101/2021.06.01.446586

Vos-Fels, K. P., Qian, L., Parra-Londono, S., Uptmoor, R., Frisch, M., and Keeble-Gagnère, G. (2017). Linkage drag constrains the roots of modern wheat. *Plant Cell Environ.* 40, 717–725. doi: 10.1111/pce.12888

Yang, Q., Lin, G., Lv, H., Wang, C., Yang, Y., and Liao, H. (2021). Environmental and genetic regulation of plant height in soybean. *BMC Plant Biol.* 21, 1–15. doi: 10.1186/s12870-021-02836-7

Yuan, X., Luo, K., Zuo, J., Liu, S., Yong, T., and Yang, W. (2022). Compensatory Growth of Soybean after Shade during Vegetative Promotes Root Nodule Recovery. *Legume Res.* 45, 1273–1277. doi: 10.18805/LRF-689

Zhang, Z., Ersoz, E., Lai, C. Q., Todhunter, R. J., Tiwari, H. K., Gore, M. A., et al. (2010). Mixed linear model approach adapted for genome-wide association studies. *Nat. Gen.* 42, 355–360. doi: 10.1038/ng.546

Zhou, W., Yao, R., Li, H., Li, S., and Yan, J. (2013). New perspective on the stabilization and degradation of the f-box protein COI1 in arabidopsis. *Plant signalling Behav.* 8 (8), e24973. doi: 10.4161/psb.24973

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 06 January 2025
DOI 10.3389/fpls.2024.1496770

Check for updates

OPEN ACCESS

EDITED BY
Mahendar Thudi,
Fort Valley State University, United States

REVIEWED BY
Venkatesh Bollina,
Agriculture and Agri-Food Canada (AAFC), Canada
Kumar Shrestha,
University of Nebraska-Lincoln, United States

*CORRESPONDENCE
Lei Zhang
zhanglei9296@aliyun.com
Yong Liu
liuyongdr@163.com
Kang Zhou
zhoukang19950601@163.com

RECEIVED 15 September 2024
ACCEPTED 02 December 2024
PUBLISHED 06 January 2025

CITATION
Deng Y, Wu W, Huang X, Yang X, Yu Y, Zhang Z, Hu Z, Zhou X, Zhou K, Liu Y and Zhang L (2025) Characterization of rhizosphere bacterial communities in oilseed rape cultivars with different susceptibility to *Plasmodiophora brassicae* infection. *Front. Plant Sci.* 15:1496770. doi: 10.3389/fpls.2024.1496770

COPYRIGHT
© 2025 Deng, Wu, Huang, Yang, Yu, Zhang, Hu, Zhou, Zhou, Liu and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Characterization of rhizosphere bacterial communities in oilseed rape cultivars with different susceptibility to *Plasmodiophora brassicae* infection

Yue Deng[1,2], Wenxian Wu[2], Xiaoqing Huang[1,2], Xiaoxiang Yang[1,2], Yaoyin Yu[1,2], Zhongmei Zhang[1,2], Zijin Hu[1], Xiquan Zhou[1], Kang Zhou[3,4]*, Yong Liu[1,2]* and Lei Zhang[1,2]*

[1]Institute of Plant Protection, Sichuan Academy of Agricultural Sciences, Chengdu, China, [2]Key Laboratory of Integrated Pest Management on Crops in Southwest, Ministry of Agriculture and Rural Affairs, Chengdu, China, [3]Anhui Province Key Laboratory of Environmental Hormone and Reproduction, Fuyang Normal University, Fuyang, China, [4]Anhui Province Key Laboratory of Embryo Development and Reproductive Regulation, Fuyang Normal University, Fuyang, China

Rhizosphere microbiomes are constantly mobilized during plant–pathogen interactions, and this, in turn, affects their interactions. However, few studies have examined the activities of rhizosphere microbiomes in plants with different susceptibilities to soil-borne pathogens, especially those that cause clubroot disease. In this study, we compared the rhizosphere bacterial community in response to infection of *Plasmodiophora brassicae* among the four different clubroot susceptibility cultivars of oilseed rape (*Brassica napus*). Our results revealed obvious differences in the responses of rhizosphere bacterial community to the *P. brassicae* infection between the four cultivars of oilseed rape. Several bacterial genera that are associated with the nitrogen cycle, including *Limnobacter*, *Thiobacillus*, *Anaeromyxobacter*, *Nitrosomonas*, *Tumebacillus*, and *Halomonas*, showed significantly different changes between susceptible and resistant cultivars in the presence of *P. brassicae* infection. Moreover, increased connectedness and robustness were exhibited in the rhizosphere bacterial community co-occurrence network in clubroot-susceptible cultivars that were infected with *P. brassicae*, while only slight changes were observed in clubroot-resistant cultivars. Metagenomic analysis of microbial metabolism also indicated differences in the rhizosphere bacterial community between susceptible and resistant cultivars that were infected with *P. brassicae*. Functional analysis of the nitrogen cycle showed that genes related to nitrification (*nxrB*) were upregulated in susceptible cultivars, while genes related to assimilatory nitrate reduction (*nasA*, *narB*, and *nirA*) were upregulated in resistant cultivars that were infected with *P. brassicae*. These findings indicate that the synthesis and assimilation process of NO_3^- content were promoted in susceptible and resistant cultivars, respectively. Our study revealed differences in

the characteristics of rhizosphere bacterial communities in response to *P. brassicae* infection between clubroot-susceptible and clubroot-resistant cultivars as well as the potential impact of these differences on the plant–*P. brassicae* interaction.

KEYWORDS

rhizosphere microbiome, *Plasmodiophora brassicae*, susceptible cultivar, resistant cultivar, microbial metabolism, nitrogen cycle

1 Introduction

Plant rhizosphere microbiomes are increasingly being studied in recent years in relation to their contributions to various aspects of plant growth, development, and health (Berendsen et al., 2012; Bardgett and van der Putten, 2014; Saleem et al., 2019). Rhizosphere microbiomes are sensitive to alterations in biotic and abiotic factors, including plant development, infection by pathogens, and disorders in soil properties, resulting in a highly dynamic and diverse microbial community throughout the entire life cycle of plants (Zegeye et al., 2019; Panke-Buisse et al., 2015; Kwak et al., 2018; Sun et al., 2020). Moreover, rhizosphere microbiomes are responses to plant/soil-borne pathogen interactions and have been widely studied in recent years. Furthermore, there is a two-way effect between microbial community assemblage and plant/pathogen interaction. In the plant/pathogen systems, plants secrete root exudates and recruit specific microbial communities that confer them with disease resistance according to the "cry for help" theory (Hu et al., 2020, 2018). Similarly, some microbes are conducive to pathogen invasion through nutritional complementarity feedback mechanisms (Li et al., 2019; Pacheco et al., 2019; Kramer et al., 2020). Changes in the microbial community of diseased plants compared with those of healthy plants are external manifestations of disease processes, and this contributes to variations in microbial metabolisms associated with energy and material metabolic cycles, such as carbon and nitrogen cycles (Cao et al., 2024). Therefore, recent studies have focused mainly on identifying certain microbial taxa in rhizosphere microbiomes that contribute to crop health by affecting microbial metabolism or inhibiting pathogen growth. These microorganisms are considered potential tools for soil-borne disease control and sustainable farming and have been widely studied in multiple crops (Sun et al., 2022; Zhang et al., 2022; Compant et al., 2019). However, relatively limited studies have focused on the impact of changes in rhizosphere microbiomes in response to plant/pathogen interactions on the development of crop disease.

Clubroot, caused by the obligate protist, *Plasmodiophora brassicae* (*P. brassicae*), is a soil-borne disease that threatens the production of Cruciferous crops worldwide, as it results in a significant reduction of 40% to 60% in both crop yield and quality (Javed et al., 2023; Chai et al., 2014). The typical symptoms of clubroot disease are the presence of root galls as well as wilting and stunting of the above-ground parts of the plant (Schuller and Ludwig-Müller, 2016; Kageyama and Asano, 2009). Traditional measures for clubroot control cannot achieve the goals of eradication due to the long survival times (5–20 years) of resting spores in the soil (Donald and Porter, 2014). Moreover, severely affected fields are unsuitable for crop cultivation for extended periods of time (Dixon, 2009). Genetic resistance is presently being considered the most economical and effective approach for clubroot control worldwide (Diederichsen et al., 2009; Rahman et al., 2014). Multiple potential clubroot resistance genes (CR gene) in *Brassica* crops that are involved in modulating disease resistance responses to *P. brassicae* infections have been identified using next-generation sequencing (NGS) and other "omics"-based methods (Nagaoka et al., 2010; Hasan et al., 2021). Furthermore, numerous clubroot-resistant cultivars of *Brassica* crops have been bred and promoted commercially, including Huashuang 5R and Huayouza 62R (Zhan et al., 2015; Shah et al., 2019). *Brassica* crops with CR loci have been shown to successfully resist *P. brassicae* infection through regulating the plant innate immunity (Zhou et al., 2020). Differences in clubroot resistance between susceptible and resistant cultivars were directly reflected in plant roots, including transcription, proteins, and metabolism (Chen et al., 2015; Zhang et al., 2016; Pedras et al., 2008; Cao et al., 2008; Li et al., 2022). However, studies on whether these CR genes participate in interactions between host root and soil microbiome during the clubroot disease process are limited.

Recent studies on clubroot have indicated a distinct shift in microbial communities of *Brassica* crops when infected with *P. brassicae* (Kang et al., 2024; Wu et al., 2020). Moreover, variations in microbial community diversity have been shown to be correlated with clubroot disease severity and are highly sensitive as indicated by the microbial community in response to infections with *P. brassicae* (Lebreton et al., 2019; Liu et al., 2024). Changes in the microbial community due to host/*P. brassicae* interactions also differ in multiple situations, including disease resistance, pathotype, soil property, and fertilization (Lebreton et al., 2019; Liu et al., 2024; Cordero-Elvia et al., 2024; Gazengel et al., 2021). A more diverse microbial community also appeared to have a more obvious effect in promoting clubroot occurrence; however, this effect varied between susceptible and resistant cultivars (Wang

et al., 2023; Daval et al., 2020). Moreover, the positive impact of nitrogen supply on clubroot occurrence also varies between susceptible and resistant cultivars (Gazengel et al., 2021). These phenomena suggest that clubroot resistance mechanisms in resistant cultivar may participate in the interaction within plants, pathogens, and soil microbiomes and may play an important role in shaping microbial communities. Although several studies have shown significant differences in root performance between clubroot-susceptible and clubroot-resistant cultivars when infected with *P. brassicae*, it is still unclear how clubroot-resistant cultivars manipulate the shaping of microbial communities based on their resistance mechanisms when infected with *P. brassicae*.

Additionally, we selected four cultivars of *Brassica napus* with different clubroot resistance levels to investigate whether clubroot resistance mechanisms affect the response of the rhizosphere microbiomes to plant/*P. brassicae* interactions. Based on 16S rRNA and metagenomic sequencing, we revealed the differences in microbial community, interaction within microbial communities, and microbial metabolisms of rhizosphere microbiomes between clubroot-susceptible and clubroot-resistant cultivars with or without *P. brassicae* infection. Our results suggest that those obvious differences in the rhizosphere microbiomes between two type cultivars may be caused by a resistant mechanism based on CR genes, which further affects the plant/*P. brassicae* interactions. Our study will help broaden the strategies for clubroot resistance breeding of oilseed rape and lay the foundation for utilizing soil microbial communities to control the occurrence of clubroot disease.

2 Materials and methods

2.1 Biological materials and pathogen inoculation

The cultivars of oilseed rape (*Brassica napus* subs. napus, hybrid) C36, H62, H62R, and Menh, as well as *P. brassicae* pathotype 4 isolate (Williams, 1966) were used in this study (Table 1). C36, also referred to as Chuanyou 36, was provided by the Crop Research Institute, Sichuan Academy of Agricultural Sciences (China) (Jiang et al., 2011). H62 and H62R, referred to as Huayouza62 (clubroot susceptible) and Huayouza62R (clubroot resistant), were provided by Prof. Chunyu Zhang of the College of Plant Science and Technology, Huazhong Agricultural University (China) (Li et al., 2021). Menh, referred to as Menhir (clubroot resistant), was provided by Norddeutsche Pflanzenzucht Hans Georg Lembke KG (NPZ) (Germany, https://www.proplanta.de/pflanzenbauberater/sorten/menhir-winterraps-hauptfruchtanbau_sks_4351raw1.html). Meanwhile, cultivars C36 and H62 are conventional hybrids without any clubroot resistance genes. H62R was generated by crossbreeding with H62 (*Brassica napus*, recipient parent) and CR Shinki (Chinese cabbage, *CRb*, donor parent). Menh was generated by crossbreeding with clubroot-resistant (*P. brassicae* pathotype 3) Mendel. Resting spores were extracted from galled root tissue collected from *Brassica napus* 'Chuanyou 81' in Shifang City, Sichuan Province. Resting spore suspensions were prepared as described in a previous study (Strelkov et al., 2006). The spore suspension was adjusted to a final concentration of 10^7 spores/mL. Each plant was inoculated with 10 mL of resting spore suspension. The resting spore suspension was injected into the soil close to the plant using a 10-mL syringe to ensure successful infection.

2.2 Experimental design and rhizosphere soil sample collection

All experiments were conducted in a greenhouse (23°C, 16-h light/8-h dark). Organic matter soil was purchased from a local market and directly used for plant cultivation without sterilization. Four cultivars of oilseed rape were also planted separately in organic matter soil in plugs (size: 540 mm × 280 mm), with each plug containing 21 holes for sowing oilseed rape. Each cultivar of oilseed rape was planted with a total of six plugs. Three plugs of plants per cultivar with a total of 63 plants were inoculated by resting spore suspensions of *P. brassicae* at 7 days; the rest of the three plugs of plants were treated with sterile water as the control treatment. For the *P. brassicae*-treated group, clubroot incidence (CI) and the disease severity index (DSI) of clubroot in each plug were calculated after 40 days of inoculation, and the rhizosphere soil of all diseased plants in the same plug were collected simultaneously into a single bag (Kuginuki et al., 1999). Regarding the control treatment, the rhizosphere soil of all plants in the same plug was also collected simultaneously into a single bag. Finally, a total of 24 rhizosphere soil samples were collected, and each treatment contained three replicates for the subsequent sequencing.

Collection and pretreatment of rhizosphere soil were done as described in a previous study (Edwards et al., 2015). The soil that remained tightly adhered to the roots after intense shaking was used as the rhizosphere soil sample. Root samples were collected into a 50-mL centrifuge tube with 25 mL of 1× PBS solution (137 mM

TABLE 1 Summary of oilseed rape cultivars used in this study.

Sample	Year	Resource	CR locus	Gene source	Reference
Chuanyou36 (C36)	2011	China	No	–	Jiang et al., 2011
Huayouza62 (H62)	2011	China	No	–	–
Huayouza62R (H62R)	2021	China	*CRb*	Gelria R	Li et al., 2021
Menhir	2015	Germany	Unknown	Unknown	North German plant breeding Hans-Georg Lembke KG

NaCl, 2.7 mM KCl, 10 mM $Na_2HPO_4{\cdot}12H_2O$, and 2 mM KH_2PO_4). The mixture was sonicated at 40 Hz for 1 min and then shaken to separate rhizosphere soil from the roots. The rhizosphere soil was then transferred to a new sterile 50-mL centrifuge tube and centrifuged at 9,000 rpm for 5 min, after which the precipitated rhizosphere soil was subjected to freeze drying (BILON-FD80AD, Shanghai Bilang Instrument Manufacturing Co., Ltd., China). The dry rhizosphere soil was then homogenized by grinding (Tissuelyser-48, Jingxin, China). Finally, the processed samples were stored at -80°C for the subsequent 16S rRNA and metagenome sequencing.

2.3 DNA extraction and 16S rRNA sequencing

Microbial DNA was extracted from 5 g of rhizosphere soil samples of oilseed rape using the E.Z.N.A.® stool DNA Kit (Omega Bio-tek, Norcross, GA, U.S.) according to the manufacturer's instructions. DNA samples were prepared and stored at -80°C for the subsequent 16S rRNA and metagenomic sequencing. The V3–V4 region of the 16S rRNA gene was PCR-amplified (95°C for 2 min, followed by 25 cycles at 95°C for 30 s, 55°C for 30 s, 72°C for 30 s, and a final extension at 72°C for 5 min) to investigate bacterial communities using the primers 338F (5′-ACTCCTACGGGAGGCAGCA-3′) and 806R (5′-GGACTACHV GGGTWTCTAAT-3′) (Mori et al., 2013). PCR reactions were performed in triplicate 20-μL mixtures containing 4 μL of 5× FastPfu buffer, 2 μL of 2.5 mM dNTPs, 0.8 μL of each primer (5 μM), 0.4 μL of FastPfu Polymerase, and 10 ng of template DNA. The PCR products were detected using 1.5% agarose gel electrophoresis and further purified using an AxyPrepTM DNA Gel Extraction Kit (Axygen Scientific, USA). PCR products were quantified using Qubit®3.0 (Life Invitrogen) and pooled in equimolar concentrations of 10 ng/μL. Paired-end sequencing was performed on an Illumina HiSeq 2500 platform at Beijing Biomarker Technologies Co., Ltd., Beijing, China. Microbial bioinformatic analysis was performed using QIIME 2 2021.11 (Bolyen et al., 2019). The raw sequencing data was demultiplexed and filtered using the q2-demux plugin followed by denoising with DADA2 (Callahan et al., 2016). The phylogenetic affiliation of each 16S rRNA gene sequence was analyzed using RDP Classifier (http://rdp.cme.msu.edu/) against the silva (SSU132) 16S rRNA database using a confidence threshold of 70% (Amato et al., 2013).

2.4 Metagenomic sequencing

Total DNA was also extracted from the above-mentioned rhizosphere soil samples using the E.Z.N.A.® Viral DNA Kit (Omega Bio-tek, Norcross, GA, USA) according to the manufacturer's protocols. High-quality DNA sample (OD260/280 = 1.8–2.2, OD260/230 ≥ 2.0) was used to construct a sequencing library. Metagenomic shotgun sequencing libraries were constructed and sequenced at Shanghai Biozeron Biological Technology Co., Ltd. Briefly, for each sample, 1 μg of genomic DNA was sheared by Covaris S220 Focused-ultrasonicator (Woburn, MA, USA), and sequencing libraries were prepared with a fragment length of approximately 450 bp. All samples were sequenced using the Illumina NovaSeq 6000 platform at Shanghai Biozeron Biotechnology Co., Ltd., Shanghai, China.

Raw sequence reads underwent quality trimming using Trimmomatic v0.36 to remove adaptor contaminants and low-quality (quality below 20 and shorter than 50 bp) reads (Bolger et al., 2014). The taxonomy of clean reads for each sample was determined by Kraken2 using the customized kraken database. The abundances of taxonomy were estimated using Bracken (https://ccb.jhu.edu/software/bracken/) which can produce accurate species- and genus-level abundance even in multiple near-identical species. Clean sequence reads were assembled with MegaHit (v1.1.1-2-g02102e1). Assembled contigs were predicted using METAProdigal (v2.6.3), and a set of unique genes were generated using CD-HIT. Gene prediction was generated using MetaGeneMark software to identify coding regions in the genome. A non-redundant genome set (95% similarity threshold, 90% coverage threshold) was conducted using MMseq2 software. GO (Gene Ontology) annotation was performed using the *goatools* package. The unique gene set was first translated into protein sequences and then searched against the NCycle database (Tu et al., 2019) using DIAMOND (Buchfink et al., 2015) to identify the gene functions with the following filter parameters: evalue 0.00001, identity 90%. CAZymes were annotated using HMMER (v.3.2.1) to match the protein sequences to entries in the hidden Markov model (HMM) libraries of CAZyme (carbohydrate-active enzymes database) families downloaded from the CAZy (Lombard et al., 2014) database (http://www.cazy.org/, v12). KEGG (Kyoto Encyclopedia of Genes and Genomes) ortholog annotation was performed using KofamScan (https://www.genome.jp/tools/kofamkoala/) with the HMMSEARCH package (https://www.ebi.ac.uk/Tools/hmmer/search/hmmsearch).

2.5 Bioinformatic analyses

The composition of the rhizosphere bacterial community at the phylum level based on the OTUs (operational taxonomic units) data was generated using the *ggplot2* package in R (v 4.3.1). The α-diversity of the bacterial community was estimated using the non-parametric Shannon and Chao1 indices. A principal coordinate analysis (PCoA) based on Bray–Curtis distance metrics was performed with R (version 4.3.1) using the *vegan* package to explore differences in bacterial community compositions between clubroot-susceptible and clubroot-resistant cultivars of oilseed rape. The Bray–Curtis distance was generated based on OUT datasets at the genus level. Multivariate analysis of variance (MANOVA) was conducted based on Bray–Curtis distance metrics to further confirm the observed differences. The heatmap of the down- and upregulation of bacterial genera in the four cultivars of oilseed rape was calculated based on the relative abundance of each bacterial genera data in all the sequencing samples. Significance analysis of bacterial genera in the four cultivars of oilseed rape was generated using the STAMP software. Commonality analysis of variation in bacterial genera

between the four cultivars was performed using an upset-venn diagram. The upset-venn diagram was completed using Wekomo Bioincloud (https://www.bioincloud.tech) (Gao et al., 2024).

Based on the disease index of clubroot and commonality analysis of variation in bacterial genera between the four cultivars, C36 and H62 were classified as clubroot-susceptible types, and H62R and Menhir were classified as clubroot-resistant types (Table 2; Figure 1F). The averages of sequencing data from two cultivars representing the clubroot-susceptible or clubroot-resistant type data were used for the subsequent analysis. The bubble diagram was completed using Wekomo Bioincloud. Microbial co-occurrence networks were used to uncover the potential interactions between rhizosphere microbiomes for clubroot-susceptible and clubroot-resistant oilseed rape cultivars with or without *P. brassicae* treatment. For each treatment, we constructed one network to display the co-occurrence patterns of bacterial ASVs (amplicon sequence variants) in the rhizosphere with or without exposure to *P. brassicae* infection. Bacterial ASVs (with a relative abundance >0.1% for at least one sample) in the rhizosphere were selected for network construction. A pairwise Spearman correlation matrix was calculated with the "corr.test" function in the *psych* package in R (version 4.3.1). Robust correlations with Spearman's correlation coefficients (ρ) > 0.6 or < -0.6 and $p < 0.01$ were used to construct networks. Network properties were performed in the *igraph* package in R (version 4.3.1). SIMPER (similarity percentages) analysis, which was completed based on the abundance data of bacterial ASVs using Wekomo Bioincloud, was used to identify the key bacterial genera that contribute significant differences in the rhizosphere bacterial co-occurrence network between susceptible and resistant cultivars. The mean relative abundance of bacterial genera was generated using the *ggplot2* package in R (version 4.3.1). Alpha and beta diversity of GO, KEGG, and CAZy pathways were performed using the *ggplot2* package in R (version 4.3.1). Significant difference analyses of the relative abundance of the KEGG pathway were completed using the STAMP software.

2.6 Data availability

The 16S rRNA amplicon data (SAMN40276299-SAMN40276322) and metagenome data (SAMN40350186-SAMN40350209) associated with this study have been deposited in the NCBI sequence read archive (SRA) under project accession PRJNA1084241. Source data have been provided in this article.

3 Results

3.1 Diversity of rhizosphere bacterial communities among the four oilseed rape cultivars infected with *P. brassicae*

The greenhouse experiment revealed that the four oilseed rape cultivars had varying susceptibility to *P. brassicae* infection. Data of CI and DSI showed that C36 and H62 cultivars were more susceptible than H62R and Menhir cultivars to *P. brassicae* infection (Table 2). We also tested rhizosphere bacterial community diversity in all treatments through 16S rRNA sequencing. A total of 794,568 16S rRNA gene reads were obtained from 16S rRNA sequencing data of all samples, with 2,277 bacterial zero-radius OTUs identified. The rhizosphere bacterial community in all samples was mainly dominated by phylum *Proteobacteria* (reads: 229,250), *Planctomycetota* (reads: 102,023), *Acidobacteriota* (reads: 96,102), *Chloroflexi* (reads: 68,968), *Actinobacteriota* (reads: 63,380), *Gemmatimonadota* (reads: 60,356), and *Bacteroidota* (reads: 31,054) (Figure 1A). The infection of *P. brassicae* did not significantly affect the rhizosphere bacterial α-diversity of oilseed rape at 40 dpi. Compared with the control group, the bacterial Shannon (6.58–6.78) and Chao1 (1,237–1,483) indices showed no significant difference in the four cultivars that were treated for *P. brassicae* ($P > 0.05$, Figure 1B). However, shifts in the rhizosphere bacterial communities were varied in the four cultivars under *P. brassicae* infection. PCoA analysis at the genus level indicated a significant difference in the structure of the rhizosphere bacterial community between C36 and Menhir cultivars in the control treatment, excluding H62 and H62R (Figure 1C). Consistently, MANOVA analysis confirmed that the cultivar type was the main driver of rhizosphere bacterial β-diversity under normal growth conditions ($R^2 = 0.58$, $P < 0.001$). However, compared with the control, the responses among the four cultivars differed when infected with *P. brassicae*, suggesting that the rhizosphere bacterial

TABLE 2 Summary of clubroot disease indices in all the treatment groups in this study.

Sample	Treatment	Number of plants	Clubroot incidence (%)	Disease severity index
C36C	Control	63	-	-
H62C	Control	63	-	-
H62RC	Control	63	-	-
MenhC	Control	63	-	-
C36T	*P. brassicae*-treated	63	90.5 ± 7.8 a	74.1 ± 3.3 a
H62T	*P. brassicae*-treated	63	79.4 ± 6.0 a	61.4 ± 6.4 b
H62RT	*P. brassicae*-treated	63	39.7 ± 9.8 b	20.1 ± 6.5 c
MenhT	*P. brassicae*-treated	63	28.6 ± 11.7 b	11.1 ± 3.9 c

Different letters (a,b,c) represent significant differences among the treatments (Tukey test. $P < 0.05$, n = 3).

10.3389/fpls.2024.1496770

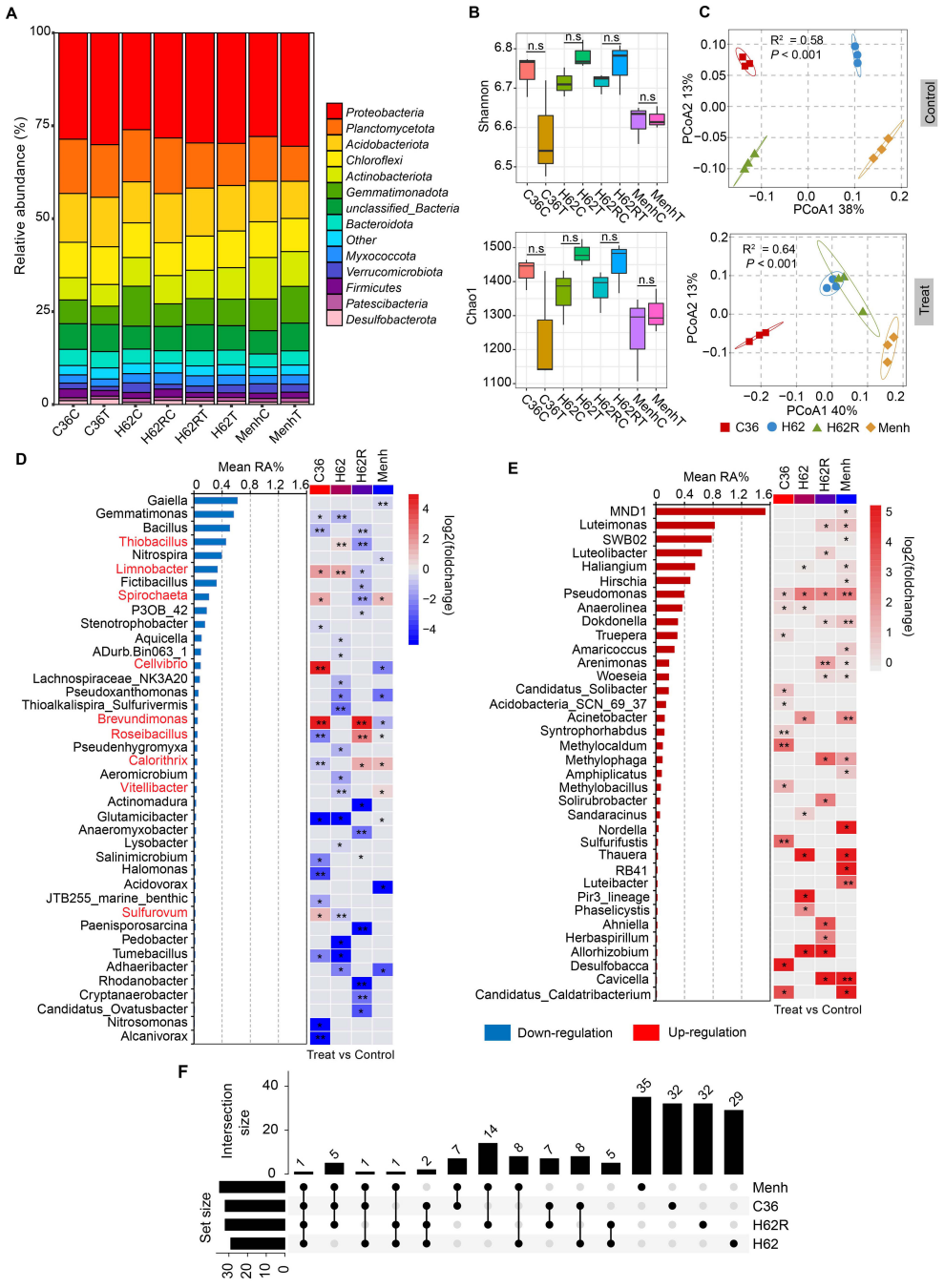


FIGURE 1
Comparison of rhizosphere bacterial diversity among the four of oilseed rape cultivars with or without *P. brassicae*. **(A)** Composition of rhizosphere bacterial community in the phylum level in all samples. C36C, H62C, H62RC, and MenhC means cultivars Chuanyou36, Huayouza62, Huayouza62R, and Menhir treated with water (Control). C36T, H62T, H62RT, and MenhT means four oilseed rape cultivars treated with resting spores of *P. brassicae*. **(B)** Rhizosphere bacterial α-diversity (Shannon and Chao1 indices) in all samples. "n.s." means no significant difference between two samples ($P > 0.05$, $n = 3$, Student's *t*-test). **(C)** PCoA of rhizosphere bacterial community in C36, H62, H62R, and Menh of oilseed rape cultivars between the *P. brassicae*-treated and control groups. **(D)** Downregulation of bacterial genera in the four cultivars with *P. brassicae*-treated group compared with those of controls. "*, **" indicate significant differences among the samples ($P < 0.05$ and $P < 0.01$, $n = 3$, Student's *t*-test). **(E)** Upregulation of bacterial genera in the four cultivars with *P. brassicae* treatment compared with the controls. "*, **" indicate significant differences among the samples ($P < 0.05$ and $P < 0.01$, $n = 3$, Student's *t*-test). **(F)** Commonality analysis of variation in bacterial genera among the four cultivars based on the statistical results shown in **(E, F)**.

community changed in response to *P. brassicae* infection among the four cultivars ($R^2 = 0.64$, $P < 0.001$, Figure 1C).

Compared to the controls, the relative abundance of 76 (76/443, 17.2%) bacterial genera was significantly changed in the four cultivars when infected with *P. brassicae* ($P < 0.05$). A relative abundance of 40 bacterial genera was also significantly downregulated in the four cultivars when infected with *P. brassicae*, wherein nine bacterial genera underwent different

changes in the four cultivars (Figure 1D). The rest of the 36 bacterial genera in the relative abundance level was significantly upregulated in the four cultivars (Figure 1E). The commonality analysis of variation in bacterial genera based on the upset-venn diagram showed that the four cultivars could be classified into two categories, C36 and H62 (top 2 intersection size: 8) and H62R and Menh (top 1 intersection size: 14) (Figure 1F)—for example, the relative abundance of *Gemmatimonas*, *Glutamicibacter*, and *Tumebacillus*, respectively, were significantly downregulated in the C36 and H62 cultivars. The relative abundance of *Limnobacter* was significantly upregulated in the C36 and H62 cultivars, while it was significantly downregulated in the H62R cultivar (Figure 1D). The relative abundance of *Luteimonas*, *Dokdonella*, *Arenimonas*, *Woeseia*, *Methylophaga*, and *Cavicella* were only significantly upregulated in the H62R and Menh cultivars (Figure 1E).

3.2 Differences in rhizosphere bacterial communities among different cultivars with varying susceptibility to infections by *P. brassicae*

Two categories, clubroot-susceptible (C36 and H62) and clubroot-resistant cultivars (H62R and Menh), were classified based on DSI and commonality analysis of variation in rhizosphere bacterial genera for subsequent analysis (Table 2; Figure 1F). Compared with the control groups, the bacterial Shannon index showed no significant differences in both clubroot-susceptible and clubroot-resistant cultivars infected with *P. brassicae* ($P > 0.05$, Figure 2A). However, the bacterial β-diversity showed significant differences among the four groups (R-C, R-T, S-C, and S-T), indicating an obvious difference in bacterial community composition between them ($R^2 = 0.15$, $P < 0.05$, Figure 2B). Compared to the control group, a total of 49 bacterial genera with relative abundance levels were significantly changed in the clubroot-susceptible ($N = 22$) or clubroot-resistant ($N = 24$) cultivars under *P. brassicae* infestation (Figure 2C). Meanwhile, the relative abundance of *Pseudomonas* and *Amaricoccus* were significantly upregulated in both susceptible and resistant cultivars. Some bacterial genera, such as *Limnobacter*, *Thiobacillus*, and *Anaeromyxobacter*, were significantly upregulated in susceptible cultivars, while thy were downregulated in resistant cultivars. Furthermore, *Nitrosomonas*, *Tumebacillus*, and *Halomonas* were significantly downregulated in susceptible cultivars; however, these were upregulated in resistant cultivars.

Moreover, a rhizosphere bacterial co-occurrence network was generated to evaluate the interaction of rhizosphere bacterial communities of oilseed rape with or without *P. brassicae* infection. Compared to the control group (susceptible control, S-C), increased connectedness and robustness were exhibited in the rhizosphere bacterial community in susceptible cultivars infected with *P. brassicae* (susceptible treat, S-T) (Figure 2D). Consistently, the number of nodes and edges in the bacterial co-occurrence network in the S-T treatment was also higher than those in the S-C treatment group (S-C: nodes = 1,444, edges = 6,611, S-T: nodes = 1,962, edges = 14,376). On the contrary, similar changes were not observed in clubroot-resistant cultivars, while the connectedness and robustness of the bacterial co-occurrence network was slightly affected by *P. brassicae* infections (Figure 2D). The number of nodes and edges in the bacterial co-occurrence network showed a similar level between R-T and R-C treatments (R-C: nodes = 1,514, edges = 10,096; R-T: nodes = 1,552, edges = 8,441), while it was lower than those in the S-T treatment (Figure 2D). The above-mentioned results implied that *P. brassicae* infection had a significant impact on the interaction of the rhizosphere bacterial community in susceptible cultivars. The SIMPER analysis identified the top 10 bacterial genera responsible for differences in microbial co-occurrence network between susceptible and resistant cultivars (Figure 2E). Meanwhile, seven bacterial genera coexist in both susceptible and resistant cultivars, with the top three contributing bacterial genera being *unclassified_Gemmatimonadaceae*, *unclassified_Pirellulaceae*, and *Pirellula*. Compared to the controls, the relative abundance of *unclassified_Gemmatimonadaceae* was upregulated in resistant cultivars when subjected to *P. brassicae* infestation, while it was significantly downregulated in susceptible cultivars ($P < 0.05$) (Figure 2F). However, the relative abundance of *unclassified_Pirellulaceae* and *Pirellula*, respectively, were downregulated in resistant cultivars, while it was upregulated in susceptible cultivars ($P > 0.05$).

3.3 Functional diversity in the rhizosphere microbial community of clubroot-susceptible and clubroot-resistant oilseed rape cultivars exposed to *P. brassicae* infestation

Metagenomic analysis was used to investigate functional differences in the rhizosphere microbial community between clubroot-susceptible and clubroot-resistant cultivars when exposed to *P. brassicae* infection. An analysis of functional diversity annotated based on GO and KEGG databases showed that there was a significant difference between susceptible and resistant cultivars with or without *P. brassicae* infestation. Compared to the controls, rhizosphere microbial functional α- and β-diversity based on GO and KEGG databases showed significant changes in susceptible cultivars exposed to *P. brassicae* infection, while there was no significant change in resistant cultivars (Figures 3A, B). In contrast to the two types previously mentioned, rhizosphere microbial functional α- and β-diversity based on CAZy database showed no significant changes in both of the two cultivars exposed to *P. brassicae* infection compared with the controls (Figure 3C).

An analysis of KEGG pathways showed that the relative abundance of multiple pathways was significantly changed in susceptible and resistant cultivars under *P. brassicae* infection compared with the controls. In susceptible cultivars, the relative abundance of 11 pathways were significantly downregulated when exposed to *P. brassicae* infection compared with the controls, while the other seven pathways were significantly upregulated (Figure 3D). In the resistant cultivars, the relative abundance of 10 pathways was significantly downregulated when exposed to *P.*

10.3389/fpls.2024.1496770

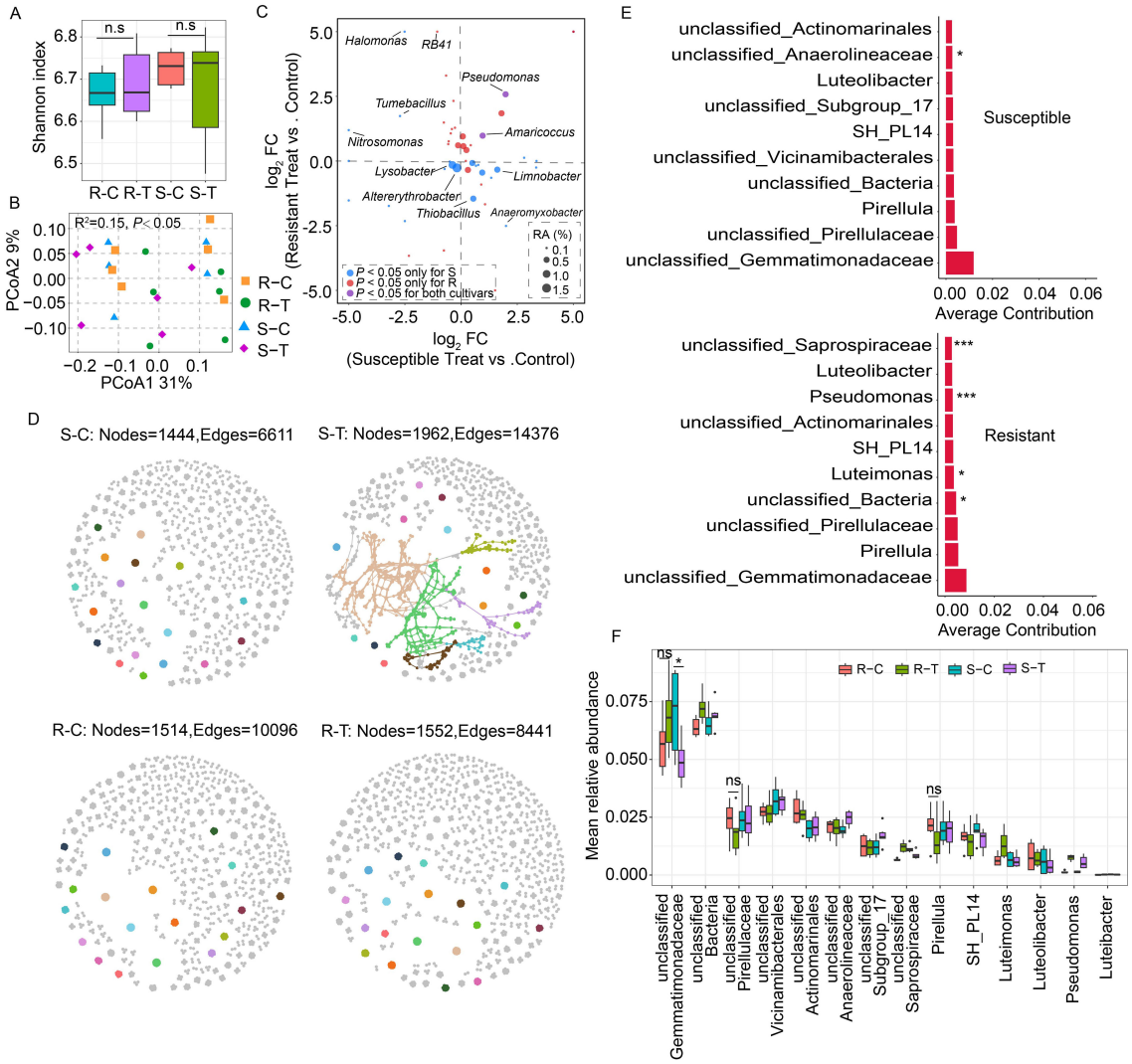


FIGURE 2
Comparison of rhizosphere bacterial community diversity between clubroot-susceptible and resistant oilseed rape cultivars. **(A)** Rhizosphere bacterial α-diversity (Shannon index) in clubroot susceptible cultivars treated with water (C36 and H62, S-C) and *P. brassicae* (S-T); clubroot-resistant cultivars treated with water (H62R and Menhir, R-C) and *P. brassicae* (R-T). "n.s." means no significant difference between two samples ($P > 0.05$, $n = 3$, Student's *t*-test). **(B)** PCoA analysis of rhizosphere bacterial community structure between susceptible and resistant cultivars. **(C)** Comparison analysis of bacterial genera that underwent significant changes in the susceptible and resistant cultivars. FC means fold change in relative abundance level between the treated and control samples. RA means the relative abundance of bacterial genus in each sample. The blue dot means bacterial genera in relative abundance levels that only showed significant changes in susceptible cultivars. The red dot means bacterial genera in relative abundance levels that only showed significant changes in resistant cultivars. The purple dot indicates bacterial genera in relative abundance level that showed significant changes in both susceptible and resistant cultivars. **(D)** Rhizosphere bacterial co-occurrence networks between clubroot-susceptible and resistant oilseed rape cultivars treated with or without *P. brassicae*. **(E)** SIMPER analysis of the top ten bacterial genera that contribute to differences in the bacterial co-occurrence network in the susceptible and resistant cultivars between the control and *P. brassicae*-treated groups. *, *** indicate significant differences among the samples ($P<0.05$, $P<0.005$, n=6, Permutation test). **(F)** Mean relative abundance of the top 10 contributing bacterial genera in the susceptible and resistant cultivar samples. "ns" means no significant difference between two samples ($P > 0.05$, $n = 6$, Student's *t*-test). "*" indicate significant differences among the samples ($P < 0.05$, n = 6, Student's t-test).

brassicae infection compared with the controls, while the other nine pathways were significantly upregulated (Figure 3D). Among them, most of the pathways belonged to microbial metabolism, and three pathways (biosynthesis of secondary metabolites, microbial metabolism in diverse environments, and nitrotoluene degradation) showed significant changes in both susceptible and resistant cultivars. However, microbial metabolism in diverse environments and nitrotoluene degradation pathways showed opposite changes between susceptible and resistant cultivars. Both of the two pathways were significantly upregulated in the susceptible cultivars under *P. brassicae* infection compared with the controls, while they were significantly downregulated in resistant cultivars.

Subsequently, we focused on the variation of genes and pathways in the nitrogen (N) cycle between susceptible and resistant cultivars with or without *P. brassicae* infection. A total of 351 unigene protein sequences (identity ≥90%) were matched to 23 genes involved in the N cycle pathway, including nitrification

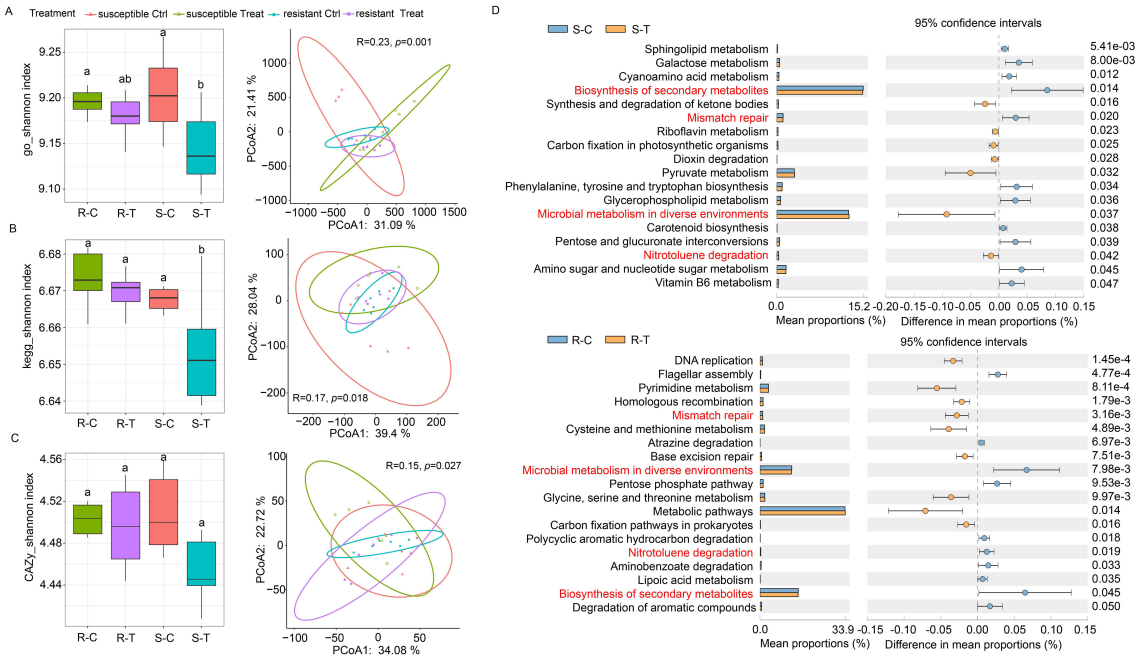


FIGURE 3
Functional diversity of rhizosphere microbiomes between clubroot-susceptible and resistant oilseed rape cultivars when exposed to *P. brassicae* infestation. **(A–C)** Functional diversity (Shannon index and PCoA) of rhizosphere microbiomes based on GO, KEGG, and CAZy databases between susceptible and resistant cultivars. Different letters represent significant differences among the treatments ($P < 0.05$, $n = 6$, Tukey test). **(D)** Differential KEGG pathways of rhizosphere microbiomes in the susceptible and resistant cultivars between the control and *P. brassicae*-treated groups.

(*amoB*, *hao*, and *nxrB*; number of unigenes = 3), denitrification (*nirS*, *nirK*, *norB*, and *nosZ*; number of unigenes = 249), nitrogen fixation (*nirH*; number of unigenes = 4), assimilatory nitrate reduction (*nasA*, *narB*, and *nirA*; number of unigenes = 8), dissimilatory nitrate reduction (*narG*, *narH*, *napA*, and *nrfA*; number of unigenes = 14), and organic nitrogen metabolism (*nmo*, gdh_K00261, gdh_K00262, gdh_K15371, *glsA*, *glnA*, *ureA*, and *ureC*; number of unigenes = 73) (Figure 4A). Among them, 16 genes were further analyzed. Compared to the controls, multiple genes showed different changes between susceptible and resistant cultivars when exposed to *P. brassicae* infection. In susceptible cultivars, we observed a higher abundance of genes associated to nitrification (*amoB* and *nxrB*), dissimilatory nitrate reduction (*narH* and *nrfA*), and denitrification (*nirS*, *nirK*, *norB*, and *nosZ*) when exposed to *P. brassicae* infection compared with the controls (Figure 4B). A variation in the *nxrB* gene abundance in the nitrification pathway ($NO_2^- \rightarrow NO_3^-$) indicated an accelerated process in NO_3^- synthesis, which was consistent with the results above (Figure 2). However, in the resistant cultivars, a number of genes associated with assimilatory nitrate reduction (*nasA*, *narB*, and *nirA*), denitrification (*nirS* and *nosZ*), nitrogen fixation (*nirH*), and nitrification (*amoB* and *hao*) were upregulated when infected with *P. brassicae* compared with the controls, while the other genes associated with denitrification (*nirK*) and organic nitrogen metabolism (*nmo*) were downregulated. Additionally, a variation in the abundance of *nasA*, *narB*, and *nirA* genes in the assimilatory nitrate reduction pathway indicated an accelerated process in NO_3^- assimilation.

4 Discussion

Alterations in the rhizosphere microbiome always occur in plants after being infected by soil-borne pathogens; this phenomenon has been reported in several crop/pathogen disease systems, including clubroot (Kwak et al., 2018; Hu et al., 2020; Ni et al., 2022). However, studies on the differences in response of rhizosphere bacterial community to *P. brassicas* infection between clubroot-susceptible and clubroot-resistant cultivars are limited. Herein we investigated the variation in rhizosphere bacterial community in oilseed rape cultivars with different susceptibility to *P. brassicae* infection. Our results revealed a distinct shift in rhizosphere bacterial communities in the four cultivars when infected with *P. brassicae* compared with the control, and this also varied among the cultivars (Figure 1). The results of this study indicated that the response of rhizosphere bacterial community to *P. brassicae* infection was different among the four cultivars, and the cultivar type was the main driving factor that led to variations in the rhizosphere bacterial community (Figure 1). Our results are consistent with those of previous studies, indicating that the host genotype has a significant impact on shaping the plant rhizosphere microbial community, which has also been extensively confirmed in other crops (Zhang et al., 2019)—for example, the Indica variety of rice showed a higher nitrogen use efficiency than that of the Japonica variety through recruiting more microbial taxa with nitrogen metabolism functions, which is determined by a nitrate transporter and sensor named *NRT1.1B*. Moreover, this distinct shift in rhizosphere bacterial community is a result of the

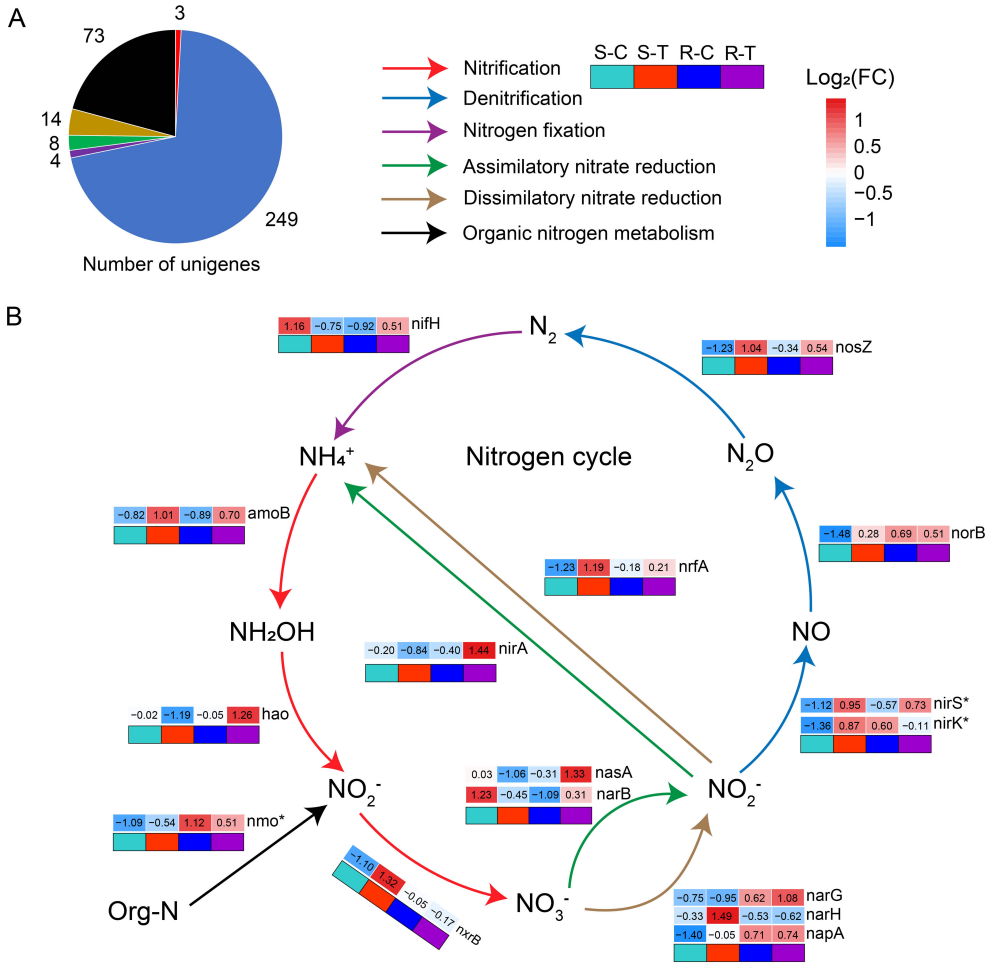


FIGURE 4
Metagenomic analysis of nitrogen cycle pathways of rhizosphere microbiomes in susceptible and resistant cultivars. **(A)** Number of unigenes annotated to genes related to the nitrogen cycle pathway. The color of the cake is consistent with that shown in **(B)**, indicating the type of nitrogen cycle pathway. **(B)** The red arrow indicates the nitrification pathway (*amoB*, *hao*, and *nxrB*) of the nitrogen cycle. The blue arrow indicates the denitrification pathway (*nirS*, *nirK*, *norB*, and *nosZ*) of the nitrogen cycle. The purple arrow indicates the nitrogen fixation pathway (*nifH*) of the nitrogen cycle. The green arrow represents the assimilatory nitrate reduction pathway (*nasA*, *narB*, and *nirA*) of the nitrogen cycle. The brown arrow indicates the dissimilatory nitrate reduction pathway (*narG*, *narH*, *napA*, and *nrfA*) of the nitrogen cycle. The black arrow represents the organic nitrogen metabolism pathway (*nmo*) of N cycle. "*" indicates significant differences among the samples ($P < 0.05$, $n = 6$, Student's *t*-test).

interaction between the host plant and *P. brassicae*. Previous studies have also proven that a distinct transcriptome landscape existed in the roots of clubroot-susceptible and clubroot-resistant Chinese cabbage lines after *P. brassicae* infection, indicating an obvious difference in clubroot resistance mechanisms and root exudations (Jia et al., 2017). A variation in the root transcription landscape could also lead to alterations in root metabolites, which, in turn, may affect the rhizosphere bacterial community (Pedras et al., 2008; Li et al., 2022). Moreover, shaping plant rhizosphere microbiomes can be achieved through the secretion of root exudates (Hu et al., 2018). In cereal crops such as wheat and maize, plant root would release benzoxazinoids to alter root-associated fungal and bacterial communities under a pathogen's infestation.

Although four cultivars showed significant differences in their rhizosphere bacterial community when infected with *P. brassicae*, clubroot occurrence performance and commonality analysis of variation in bacterial genera also confirmed that cultivars C36 and H62, H62R, and Menhir, could be classified into clubroot-susceptible and clubroot-resistant types (Table 2; Figure 1). A comparative analysis further revealed differences in rhizosphere bacterial communities between clubroot-susceptible and clubroot-resistant cultivars in response to *P. brassicae* infection (Figure 2). Compared to the controls, some bacterial genera, such as *Limnobacter*, *Thiobacillus*, *Anaeromyxobacter Nitrosomonas*, *Tumebacillus*, and *Halomonas*, showed significant changes in relative abundance level between susceptible and resistant cultivars when exposed to *P. brassicae* infection. Meanwhile, *Limnobacter*, *Thiobacillus*, and *Anaeromyxobacter* were reported to be associated with BNF (biological N_2 fixation) and were significantly upregulated in susceptible cultivars infected with *P. brassicae* and were downregulated in resistant cultivars. Additionally, *Thiobacillus* and *Anaeromyxobacter* were reported to be associated with arsenite oxidation-dependent biological nitrogen fixation, and *Limnobacter* was reported to be associated with nitrification-anammox (PN/A)

processes (Li et al., 2023; Wang P. et al., 2024). Our results suggested that variations in the relative abundance of those three bacterial genera may lead to an accumulation of NO_3^- and SO_4^{2-} in the soil and a reduction in pH value, which may be conducive to *P. brassicae* infection by reducing the soil's pH value (Wang et al., 2023). Moreover, *Nitrosomonas*, *Tumebacillus*, and *Halomonas* were significantly downregulated in susceptible cultivars but were upregulated in resistant cultivars. *Nitrosomonas*, *Tumebacillus*, and *Halomonas* were reported to be associated with nitrosation and aerobic denitrification; variations in the relative abundance of those bacterial genera have also been shown to lead to an accumulation of NO_3^- and SO_4^{2-} in soil samples from susceptible cultivars (Arp et al., 2002; Zhang et al., 2014; González-Domenech et al., 2010).

A variation in bacterial taxa is also reflected in microbial interactions. The bacterial co-occurrence network also showed a distinct interaction intensity within rhizosphere bacterial communities in susceptible cultivars infected with *P. brassicae* compared with the controls. Meanwhile, increased connectedness and robustness of the bacterial co-occurrence network were observed in susceptible cultivars exposed to *P. brassicae* infestation compared with the controls, while there were slight changes in resistant cultivars. The SIMPER analysis revealed that *unclassified_Gemmatimonadaceae*, *unclassified_Pirellulaceae*, and *Pirellula* were the three most common bacterial genera that contributed to the differences in bacterial co-occurrence network between the *P. brassicae*-treated and control samples in susceptible and resistant cultivars (Figure 2E). Among them, the relative abundance of *unclassified_Gemmatimonadaceae* was upregulated in resistant cultivars exposed to *P. brassicae*, while it was significantly downregulated in susceptible cultivars (Figure 2F). *Gemmatimonadaceae* was reported to be associated with plant root metabolites, which had a negative correlation with organic acid and a positive correlation with ketone content (Wang M. et al., 2024). Flavonoids, a type of ketone, have been reported to be associated with clubroot disease resistance in *Arabidopsis thaliana* (Päsold et al., 2010). The relative abundance of *unclassified_Pirellulaceae* and *Pirellula* was downregulated in the resistant cultivars when exposed to *P. brassicae*, while it showed slight changes in susceptible cultivars. *Pirellula* belonged to *Planctomycetes* and has been shown to be associated with anaerobic ammonia oxidation (Huang et al., 2014). In this study, a decrease in the relative abundance of *Pirellula* in resistant cultivars may also lead to the degradation of NO_3^- in the soil.

Variations in bacterial taxa and interaction within rhizosphere bacterial communities may also be reflected in microbial metabolisms. Functional diversity analysis based on metagenomic sequencing data further confirmed that changes in bacterial communities resulted in alterations in microbial metabolism. Our results showed that multiple pathways associated with microbial metabolism were significantly different in susceptible and resistant cultivars when exposed to *P. brassicae* infection compared with the controls (Figure 3). Among them, pathways associated with the nitrogen cycle exhibited distinct differences between susceptible and resistant cultivars (Figure 3D). Our results also demonstrated that the synthesis (nitrification) and assimilation (assimilatory nitrate reduction) processes of NO_3^- content were promoted in susceptible and resistant cultivars, respectively. Meanwhile, the abundance of *nxrB* gene related to nitrification (NO_2^-"NO_3^-) was upregulated in susceptible cultivars exposed to *P. brassicae* infection compared with the control, while it only changed slightly in resistant cultivars. Moreover, the expression of *nasA*, *narB*, and *nirA* genes, which are related to assimilatory nitrate reduction, was upregulated in resistant cultivars when exposed to *P. brassicae* infections compared with the controls, while it was downregulated in susceptible cultivars (Figure 4). In this study, the differences in nitrification and assimilatory nitrate reduction pathways between susceptible and resistant cultivars were due to rhizosphere microbiomes in response to *B. napus*/*P. brassicae* interaction. Our results suggested that NO_3^- may be one of the critical factors that affect *B. napus*/*P. brassicae* interaction and could reduce the incidence of clubroot.

Nitrogen, which is widely considered as a central element in soil ecosystems, has a huge impact on plant/pathogen interactions (Cui et al., 2014). Meanwhile, nitrogen supply could enhance the development of biotrophic pathogens, while the opposite effect is observed for necrotrophic pathogens (Solomon et al., 2003; Mur et al., 2016). The different forms of nitrogen supply (ammonium NH_4^+ or nitrate NO_3^-) can have various effects on the occurrence of plant disease due to differences in assimilation and metabolism pathways (Mur et al., 2016). NO_3^- feeding can strengthen host hypersensitive response (HR)-mediated resistance through enhancing the production of polyamines, while NH_4^+ nutrition can attenuate host defense. Regarding clubroot, although some studies suggest that the occurrence of clubroot is reduced with the application of high-nitrogen fertilizers, this may be attributed to the fact that oilseed rape requires a relatively large amount of nitrogen fertilizer during its entire growth period for growth and disease resistance (Gossen et al., 2014; Rathke et al., 2005). However, recent studies have shown that a high nitrogen supply could promote the occurrence of clubroot in susceptible *B. napus* cultivars by regulating the transcriptomic profile of *P. brassicae*, including pathogenicity-related genes (NUDIX and NEP-proteins) and genes associated to obligate biotrophic functions (glutamine synthetase, associated with nitrogen metabolism), whereas the effect differs in resistant cultivars (Gazengel et al., 2021). The above-mentioned results suggest that nitrogen supply may be beneficial for *P. brassicae* infection. Germination of *P. brassicae*'s resting spores is crucial for the occurrence of clubroot disease. Recent studies have proven that a diverse bacterial community, rather than root exudates, is necessary to stimulate the germination of the resting spores of *P. brassicae* (Wang et al., 2023). Meanwhile, the relative abundance of *Sphingobacteriia*, *Flavobacteriia*, and *Bacteroidetes* was significantly enriched in the "high"-germination-rate group, while *Proteobacteria* dominated in the "low"-germination-rate group. Moreover, the addition of NO_3^-, not NH_4^+, was conducive for the induction of the microbial community, leading to the germination of resting spores. The NO_3^- supply may be utilized as nutrients by certain microorganisms and could enhance nitrogen cycle pathways within microbial communities. The results of this study are consistent with the conclusions of previous studies. The NO_3^- synthesis pathways in the rhizosphere microbiomes were promoted in susceptible cultivars when exposed to *P. brassicae* infection compared with the controls, while the NO_3^- assimilation pathways in the rhizosphere microbiomes were promoted in resistant cultivars. Furthermore, in this study, variations in the NO_3^- assimilation and synthesis pathways in rhizosphere microbiomes were a result of the occurrence of clubroot, suggesting that changes in rhizosphere

microbial community were directed by the *B. napus*/*P. brassicae* interaction. We considered that this rhizosphere microbial ecology environment associated with NO_3^- accumulation in the susceptible cultivar may be conductive to the further development of clubroot. In the clubroot-susceptible cultivar, once the host's own defense system was breached, *P. brassicae* may control the rhizosphere microbial community to facilitate further infection by regulating the host's metabolism. However, in clubroot-resistant varieties, the rhizosphere microbial community may be continuously manipulated by the host to jointly resist the *P. brassicae* infection. The differences between susceptible and resistant cultivars may be determined by their distinct resistance mechanisms. Multiple studies have also shown that *Brassica* crop roots showed different changes in genes, transcription, metabolomics, and proteome perspectives after *P. brassicae* infection, while the situation varied between susceptible and resistant cultivars due to the CR gene (Chen et al., 2015; Zhang et al., 2016; Pedras et al., 2008; Cao et al., 2008; Li et al., 2022). In this study, although we revealed the differences in NO_3^- assimilation and synthesis pathways in the rhizosphere microbiomes between susceptible and resistant cultivars after infection with *P. brassicae*, it is still unclear that soil NO_3^- participated in the interaction between hosts, *P. brassicae*, and rhizosphere microbiomes. In this study, the results of microbial diversity were conducted based on relative abundance indicators, and we only focused on limited microbial taxa that were differentially distributed among different samples. Hence, we might ignore those microbial taxa that showed significant changes in absolute abundance level across different samples while having similar relative abundance levels, and these microbes might be crucial for regulating microbial ecological functions. Moreover, all the conclusions obtained in this study were completed in the greenhouse condition, so several environmental impact factors, such as soil property, fertilizer regime, and cultivation pattern, were overlooked compared to field experiments. Our further study will validate the effect of NO_3^- pathway of microbial community on clubroot occurrence and clarify the relationship between soil NO_3^- content and microbial community function. We will also investigate the succession pattern of soil nitrogen cycling during the occurrence of the clubroot disease process, which is conducive to clarify the occurrence mechanism of clubroot disease and lay a theoretical foundation for clubroot disease control by using soil microorganisms.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/, PRJNA1084241.

Author contributions

YD: Data curation, Formal analysis, Investigation, Resources, Software, Writing – original draft, Writing – review & editing. WW: Conceptualization, Investigation, Resources, Visualization, Writing – review & editing. XH: Conceptualization, Funding acquisition, Supervision, Validation, Visualization, Writing – review & editing. XY: Conceptualization, Methodology, Supervision, Writing – review & editing. YY: Investigation, Supervision, Validation, Writing – review & editing. ZZ: Conceptualization, Investigation, Supervision, Validation, Writing – review & editing. ZH: Software, Supervision, Validation, Writing – review & editing. XZ: Investigation, Writing – review & editing. KZ: Conceptualization, Formal Analysis, Software, Supervision, Validation, Visualization, Writing – review & editing. YL: Conceptualization, Data curation, Funding acquisition, Investigation, Validation, Visualization, Writing – review & editing. LZ: Conceptualization, Data curation, Funding acquisition, Project administration, Supervision, Visualization, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported financially by the Natural Science Foundation of Sichuan Province (2022NSFSC0168 and 2023NSFSC0150). This work was also supported financially by the Independent Innovation Project (2022ZZCX023), the Science and Technology Project (1 + 9KJGG006), and the Talent Research Fund (NKYRCZX2024018) of Sichuan Academy of Agricultural Sciences.

Conflict of interest

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

Publisher's note

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

References

Amato, K. R., Yeoman, C. J., Kent, A., Righini, N., Carbonero, F., Estrada, A., et al. (2013). Habitat degradation impacts black howler monkey (Alouatta pigra) gastrointestinal microbiomes. *ISME J.* 7, 1344–1353. doi: 10.1038/ismej.2013.16

Arp, D. J., Sayavedra-Soto, L. A., and Hommes, N. G. (2002). Molecular biology and biochemistry of ammonia oxidation by *Nitrosomonas europaea*. *Arch. Microbiol.* 178, 250–255. doi: 10.1007/s00203-002-0452-0

Bardgett, R. D., and van der Putten, W. H. (2014). Belowground biodiversity and ecosystem functioning. *Nature.* 515, 505–511. doi: 10.1038/nature13855

Berendsen, R. L., Pieterse, C. M. J., and Bakker, P. A. (2012). The rhizosphere microbiome and plant health. *Trends Plant Sci.* 17, 478–486. doi: 10.1016/S2095-3119(21)63817-0

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Oxford, England. *Bioinformatics.* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., et al. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* 37, 852–857. doi: 10.1038/s41587-019-0209-9

Buchfink, B., Xie, C., and Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12, 59–60. doi: 10.1038/nmeth.3176

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J., and Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* 13, 581–583. doi: 10.1038/nmeth.3869

Cao, T., Luo, Y., Shi, M., Tian, X., and Kuzyakov, Y. (2024). Microbial interactions for nutrient acquisition in soil: Miners, scavengers, and carriers. *Soil Biol. Biochem.* 188, 109215. doi: 10.1016/j.soilbio.2023.109215

Cao, T., Srivastava, S., Rahman, M. H., Kav, N. N. V., Hotte, N., Deyholos, M. K., et al. (2008). Proteome-level changes in the roots of *Brassica napus* as a result of *Plasmodiophora brassicae* infection. *Plant Sci.* 174, 97–115. doi: 10.1016/j.plantsci.2007.10.002

Chai, A. L., Xie, X. W., Shi, Y. X., and Li, B. J. (2014). Special issue: research status of clubroot (*Plasmodiophora brassicae*) on cruciferous crops in China. *Can. J. Plant Pathol.* 36, 142–153. doi: 10.1080/07060661.2013.868829

Chen, J., Pang, W., Chen, B., Zhang, C., and Piao, Z. (2015). Transcriptome analysis of *Brassica rapa* near-isogenic lines carrying clubroot-resistant and -susceptible alleles in response to *Plasmodiophora brassicae* during early infection. *Front. Plant Sci.* 6. doi: 10.3389/fpls.2015.01183

Compant, S., Samad, A., Faist, H., and Sessitsch, A. (2019). A review on the plant microbiome: Ecology, functions, and emerging trends in microbial application. *J. Adv. Res.* 19, 29–37. doi: 10.1016/j.jare.2019.03.004

Cordero-Elvia, J., Galindo-González, L., Fredua-Agyeman, R., Hwang, S.-F., and Strelkov, S. E. (2024). Clubroot-induced changes in the root and rhizosphere microbiome of susceptible and resistant canola. *Plants.* 13, 1880. doi: 10.3390/plants13131880

Cui, X., Yan, Q. W., Sun, J. L., Xiao, S., Xie, F. C., and Chen, Y. J. (2014). Research progress on nitrogen use and plant growth. *J. Northeast Agric. Univ.* 21, 68–74. doi: 10.1016/S1006-8104(14)60036-2

Daval, S., Gazengel, K., Belcour, A., Linglin, J., Guillerm-Erckelboudt, A. Y., Sarniguet, A., et al. (2020). Soil microbiota influences clubroot disease by modulating *Plasmodiophora brassicae* and *Brassica napus* transcriptomes. *Microb. Biotechnol.* 13, 1648–1672. doi: 10.1111/1751-7915.13634

Diederichsen, E., Frauen, M., Linders, E. G. A., Hatakeyama, K., and Hirai, M. (2009). Status and perspectives of clubroot resistance breeding in crucifer crops. *J. Plant Growth Regul.* 28, 265–281. doi: 10.1007/s00344-009-9100-0

Dixon, G. R. (2009). The occurrence and economic impact of *Plasmodiophora brassicae* and clubroot disease. *J. Plant Growth Regul.* 28, 194–202. doi: 10.1007/s00344-009-9090-y

Donald, E. C., and Porter, I. J. (2014). Clubroot in Australia: The history and impact of *Plasmodiophora brassicae* in *Brassica* crops and research efforts directed towards its control. *Can. J. Plant Pathol.* 36, 66–84. doi: 10.1080/07060661.2013.873482

Edwards, J., Johnson, C., Santos-Medellín, C., Lurie, E., Podishetty, N. K., Bhatnagar, S., et al. (2015). Structure, variation, and assembly of the root-associated microbiomes of rice. *Proc. Natl. Acad. Sci. U.S.A.* 112, E911–E920. doi: 10.1073/pnas.1414592112

Gao, Y. Y., Zhang, G. X., Jiang, S. Y., and Liu, Y. X. (2024). Wekemo Bioincloud: A user-friendly platform for meta-omics data analyses. *IMeta.* 3, e175. doi: 10.1002/imt2.175

Gazengel, K., Aigu, Y., Lariagon, C., Humeau, M., Gravot, A., Manzanares-Dauleux, M. J., et al. (2021). Nitrogen supply and host-plant genotype modulate the transcriptomic profile of *Plasmodiophora brassicae*. *Front. Microbiol.* 12. doi: 10.3389/fmicb.2021.701067

González-Domenech, C. M., Martínez-Checa, F., Béjar, V., and Quesada, E. (2010). Denitrification as an important taxonomic marker within the genus *Halomonas*. *Syst. Appl. Microbiol.* 33, 85–93. doi: 10.1016/j.syapm.2009.12.001

Gossen, B. D., Deora, A., Peng, G., Hwang, S.-F., and Mcdonald, M. R. (2014). Effect of environmental parameters on clubroot development and the risk of pathogen spread. *Can. J. Plant Pathol.* 36, 37–48. doi: 10.1080/07060661.2013.859635

Hasan, J., Megha, S., and Rahman, H. (2021). Clubroot in *Brassica*: recent advances in genomics, breeding, and disease management. *Genome.* 64, 735–760. doi: 10.1139/gen-2020-0089

Hu, J., Wei, Z., Kowalchuk, G. A., Xu, Y., Shen, Q., and Jousset, A. (2020). Rhizosphere microbiome functional diversity and pathogen invasion resistance build up during plant development. *Environ. Microbiol.* 22, 5005–5018. doi: 10.1111/1462-2920.15097

Hu, L., Robert, C. A. M., Cadot, S., Zhang, X., Ye, M., Li, B., et al. (2018). Root exudate metabolites drive plant-soil feedbacks on growth and defense by shaping the rhizosphere microbiota. *Nat. Commun.* 9, 2738. doi: 10.1038/s41467-018-05122-7

Huang, P. B., Jiao, N. Z., Feng, J., and Shu, Q. L. (2014). Research progress on *Planctomycetes*' diversity and ecological function in marine environments. *Microbiol. China.* 41, 1891–1902. doi: 10.13344/j.microbiol.China.130874

Javed, M. A., Schwelm, A., Zamani-Noor, N., Salih, R., Silvestre Vañó, M., Wu, J., et al. (2023). The clubroot pathogen *Plasmodiophora brassicae*: A profile update. *Mol. Plant Pathol.* 24, 89–106. doi: 10.1111/mpp.13283

Jia, H., Wei, X., Yang, Y., Yuan, Y. X., Wei, F., Zhao, Y. Y., et al. (2017). Root RNA-seq analysis reveals a distinct transcriptome landscape between clubroot-susceptible and clubroot-resistant Chinese cabbage lines after *P. brassicae* infection. *Plant Soil.* 421, 93–105. doi: 10.1007/s11104-017-3432-5

Jiang, L. C., Huang, C., Pu, X. B., Zhang, J. F., Li, H. J., Jiang, J., et al. (2011). Breeding of new high-yield, extensive-adaptability and double-low NEA-cytoplasmic hybrid variety Chuanyou No. 36 in *Brassica napus* L. (in Chinese). *Southwest China J. Agri Sci.* 24, 1660–1664. doi: 10.16213/j.cnki.scjas.2011.05.032

Kageyama, K., and Asano, T. (2009). Life cycle of *Plasmodiophora brassicae*. *J. Plant Growth Regul.* 28, 203–211. doi: 10.1007/s00344-009-9101-z

Kang, H. J., Chai, A. L., Lin, Z. L., Shi, Y. X., Xie, X. W., Li, L., et al. (2024). Deciphering differences in microbial community diversity between clubroot-diseased and healthy soils. *Microorganisms.* 12, 251–251. doi: 10.3390/microorganisms12020251

Kramer, J., Özkaya, Ö., and Kümmerli, R. (2020). Bacterial siderophores in community and host interactions. *Nat. Rev. Microbiol.* 18, 152–163. doi: 10.1038/s41579-019-0284-4

Kuginuki, Y., Yoshikawa, H., and Hirai, M. (1999). Variation in virulence of *Plasmodiophora brassicae* in Japan tested with clubroot-resistant cultivars of chinese cabbage (*Brassica rapa* L. ssp. *pekinensis*). *Eur. J. Plant Pathol.* 105, 327–332. doi: 10.1023/A:1008705413127

Kwak, M.-J., Kong, H. G., Choi, K., Kwon, S.-K., Song, J. Y., Lee, J., et al. (2018). Rhizosphere microbiome structure alters to enable wilt resistance in tomato. *Nat. Biotechnol.* 36, 1100–1109. doi: 10.1038/nbt.4232

Lebreton, L., Guillerm-Erckelboudt, A. Y., Gazengel, K., Linglin, J., Ourry, M., Glory, P., et al. (2019). Temporal dynamics of bacterial and fungal communities during the infection of *Brassica rapa* roots by the protist *Plasmodiophora brassicae*. *PloS One* 14, e0204195. doi: 10.1371/journal.pone.0204195

Li, J., Huang, T., Lu, J., Xu, X., and Zhang, W. (2022). Metabonomic profiling of clubroot-susceptible and clubroot-resistant radish and the assessment of disease-resistant metabolites. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.1037633

Li, M., Wei, Z., Wang, J., Jousset, A., Friman, V. P., Xu, Y., et al. (2019). Facilitation promotes invasions in plant-associated microbial communities. *Ecol. Lett.* 22, 149–158. doi: 10.1111/ele.13177

Li, Q., Nadil, S., Zhou, Y. W., Hou, Z. K., Gong, J. F., Liu, J., et al. (2021). Breeding of a novel clubroot disease-resistant *Brassica napus* variety Huayouza 62R. *Acta Agronomica Sin.* 47, 210–223. doi: 10.3724/SP.J.1006.2021.04086

Li, Y., Guo, L., Yang, R., Yang, Z., Zhang, H., Li, Q., et al. (2023). *Thiobacillus* spp. and Anaeromyxobacter spp. mediate arsenite oxidation-dependent biological nitrogen fixation in two contrasting types of arsenic-contaminated soils. *J. Hazard Mater.* 443, 130220. doi: 10.1016/j.jhazmat.2022.130220

Liu, Y., Lai, J., Sun, X., Huang, L., Sheng, Y., Zhang, Q., et al. (2024). Comparative metagenomic analysis reveals rhizosphere microbiome assembly and functional adaptation changes caused by clubroot disease in Chinese Cabbage. *Microorganisms.* 12, 1370. doi: 10.3390/microorganisms12071370

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., and Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* 42, D490–D495. doi: 10.1093/nar/gkt1178

Mori, H., Maruyama, F., Kato, H., Toyoda, A., Dozono, A., Ohtsubo, Y., et al. (2013). Design and experimental application of a novel non-degenerate universal primer set that amplifies prokaryotic 16S rRNA genes with a low possibility to amplify eukaryotic rRNA genes. *DNA Res.* 21, 217–227. doi: 10.1093/dnares/dst052

Mur, L. A. J., Simpson, C., Kumari, A., Gupta, A. K., and Gupta, K. J. (2016). Moving nitrogen to the centre of plant defence against pathogens. *Ann. Bot.* 119, 703–709. doi: 10.1093/aob/mcw179

Nagaoka, T., Doullah, M. A. U., Matsumoto, S., Kawasaki, S., Ishikawa, T., Hori, H., et al. (2010). Identification of QTLs that control clubroot resistance in *Brassica oleracea* and comparative analysis of clubroot resistance genes between *B. rapa* and *B. oleracea*. *Theor. Appl. Genet.* 120, 1335–1346. doi: 10.1007/s00122-010-1259-z

Ni, H., Zong, R., Sun, J., Wu, Y., Yu, L., Liu, Y., et al. (2022). Response of bacterial community to the occurrence of clubroot disease in Chinese cabbage. *Front. Microbiol.* 13. doi: 10.3389/fmicb.2022.922660

Pacheco, A. R., Moel, M., and Segrè, D. (2019). Costless metabolic secretions as drivers of interspecies interactions in microbial ecosystems. *Nat. Commun.* 10, 103. doi: 10.1038/s41467-018-07946-9

Panke-Buisse, K., Poole, A. C., Goodrich, J. K., Ley, R. E., and Kao-Kniffin, J. (2015). Selection on soil microbiomes reveals reproducible impacts on plant function. *ISME J.* 9, 980–989. doi: 10.1038/ismej.2014.196

Päsold, S., Siegel, I., Seidel, C., and Ludwig-Müller, J. (2010). Flavonoid accumulation in *Arabidopsis thaliana* root galls caused by the obligate biotrophic pathogen

Plasmodiophora brassicae. *Mol. Plant Pathol.* 11, 545–562. doi: 10.1111/j.1364-3703.2010.00628.x

Pedras, M. S. C., Zheng, Q. A., and Strelkov, S. (2008). Metabolic changes in roots of the oilseed canola infected with the biotroph *Plasmodiophora brassicae*: phytoalexins and phytoanticipins. *J. Agric. Food Chem.* 56, 9949–9961. doi: 10.1021/jf802192f

Rahman, H., Peng, G., Yu, F., Falk, K. C., Kulkarni, M., and Selvaraj, G. (2014). Genetics and breeding for clubroot resistance in canadian spring canola (*Brassica napus* L.). *Can. J. Plant Pathol.* 36, 122–134. doi: 10.1080/07060661.2013.862571

Rathke, G. W., Christen, O., and Diepenbrock, W. (2005). Effects of nitrogen source and rate on productivity and quality of winter oilseed rape (*Brassica napus* L.) grown in different crop rotations. *Field Crops Res.* 94, 103–113. doi: 10.1016/j.fcr.2004.11.010

Saleem, M., Hu, J., and Jousset, A. (2019). More than the sum of its parts: microbiome biodiversity as a driver of plant growth and soil health. *Annu. Rev. Ecol. Evol. S.* 50, 145–168. doi: 10.1146/annurev-ecolsys-110617-062605

Schuller, A., and Ludwig-Müller, J. (2016). Histological methods to detect the clubroot pathogen *Plasmodiophora brassicae* during its complex life cycle. *Plant Pathol.* 65, 1223–1237. doi: 10.1111/ppa.12520

Shah, N., Sun, J., Yu, S., Yang, Z., Wang, Z., Huang, F., et al. (2019). Genetic variation analysis of field isolates of clubroot and their responses to *Brassica napus* lines containing resistant genes *CRb* and *PbBa8.1* and their combination in homozygous and heterozygous state. *Mol. Breeding.* 39, 1–11. doi: 10.1007/s11032-019-1075-3

Solomon, P. S., Tan, K. C., and Oliver, R. P. (2003). The nutrient supply of pathogenic fungi; a fertile field for study. *Mol. Plant Pathol.* 4, 203–210. doi: 10.1046/j.1364-3703.2003.00161.x

Strelkov, S. E., Tewari, J. P., and Smith-Degenhardt, E. (2006). Characterization of *Plasmodiophora brassicae* populations from Alberta, Canada. *Can. J. Plant Pathol.* 28, 467–474. doi: 10.1080/07060660609507321

Sun, Q., Wu, H. L., Chen, F., and Kang, J. H. (2020). Characteristics of soil nutrients and fungal community composition in crop rhizosphere under different rotation patterns. *Environ. Sci.* 41, 4682–4689. doi: 10.13227/j.hjkx.202001031

Sun, X., Xu, Z., Xie, J., Hesselberg-Thomsen, V., Tan, T., Zheng, D., et al. (2022). *Bacillus velezensis* stimulates resident rhizosphere *Pseudomonas stutzeri* for plant health through metabolic interactions. *ISME J.* 16, 774–787. doi: 10.1038/s41396-021-01125-3

Tu, Q., Lin, L., Cheng, L., Deng, Y., and He, Z. (2019). NCycDB: a curated integrative database for fast and accurate metagenomic profiling of nitrogen cycling genes. *Bioinformatics.* 35, 1040–1048. doi: 10.1093/bioinformatics/bty741

Wang, M., Mu, C., Lin, X., Ma, W., Wu, H., Si, D., et al. (2024). Foliar application of nanoparticles reduced cadmium content in wheat (*Triticum aestivum* L.) grains via long-distance "leaf-root-microorganism" regulation. *Environ. Sci. Technol.* 58, 6900–6912. doi: 10.1021/acs.est.3c10506

Wang, P., Ou, R., Tan, J., Li, N., Zheng, M., Jin, Q., et al. (2024). Effect of sludge redistribution strategy on stability of partial nitrification-anammox process: Further exploration of the potential value of sludge. *Chemosphere.* 355, 141707. doi: 10.1016/j.chemosphere.2024.141707

Wang, Y., Zheng, X., Sarenqimuge, S., and von Tiedemann, A. (2023). The soil bacterial community regulates germination of *Plasmodiophora brassicae* resting spores rather than root exudates. *PloS Pathog.* 19, e1011175. doi: 10.1371/journal.ppat.1011175

Williams, P. H. (1966). A system for the determination of races of *Plasmodiophora brassicae* that infect cabbage and rutabaga. *Phytopathology.* 56, 624–626.

Wu, W. X., Huang, X. Q., Zhang, L., Yang, X. X., Li, H. Z., and Liu, Y. (2020). Crucifer clubroot disease changes the microbial community structure of rhizosphere soil. *Acta Ecologica Sinica.* 40, 1532–1541. doi: 10.5846/stxb201901190151

Zegeye, E. K., Brislawn, C. J., Farris, Y., Fansler, S. J., Hofmockel, K. S., Jansson, J. K., et al. (2019). Selection, succession, and stabilization of soil microbial consortia. *mSystems.* 4, e00055–e00019. doi: 10.1128/mSystems.00055-19

Zhan, Z. X., Jiang, Y. F., Zhu, Z. Y., Zhang, C. S., Yang, Q. Y., Li, Q., et al. (2015). Development of close linked marker to *PbBa8.1* con-ferring canola resistance to *Plasmodiophora brassicae*. *Chin. J. Oil Crop Sci.* 37, 772–779. doi: 10.7505/j.issn.1007-9084.2015.06.005

Zhang, F. T., Huang, H. Q., Cui, Y., Sun, Q. G., Zhu, J., Liu, M., et al. (2014). Isolation and diversity of bacillus species from Jiaxi tropical rain forest soil. *J. Microbiol.* 34, 42–46. doi: 10.3969/j.issn.1005-7021.2014.04.008

Zhang, J., Ahmed, W., Dai, Z., Zhou, X., He, Z., Wei, L., et al. (2022). Microbial consortia: an engineering tool to suppress clubroot of chinese cabbage by changing the rhizosphere bacterial community composition. *Biology.* 11, 918. doi: 10.3390/biology11060918

Zhang, J., Liu, Y. X., Zhang, N., Hu, B., Jin, T., Xu, H., et al. (2019). NRT1.1B is associated with root microbiota composition and nitrogen use in field-grown rice. *Nat. Biotechnol.* 37, 676–684. doi: 10.1038/s41587-019-0104-4

Zhang, X., Liu, Y., Fang, Z., Li, Z., Yang, L., Zhuang, M., et al. (2016). Comparative transcriptome analysis between broccoli (*Brassica oleracea* var. *italica*) and wild cabbage (*Brassica macrocarpa* Guss.) in response to *Plasmodiophora brassicae* during different infection stages. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.01929

Zhou, Q., Galindo-González, L., Manolii, V., Hwang, S. F., and Strelkov, S. E. (2020). Comparative transcriptome analysis of rutabaga (*Brassica napus*) cultivars indicates activation of salicylic acid and ethylene-mediated defenses in response to *Plasmodiophora brassicae*. *Int. J. Mol. Sci.* 21, 8381. doi: 10.3390/ijms21218381

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 09 January 2025
DOI 10.3389/fpls.2024.1506139

Check for updates

OPEN ACCESS

EDITED BY
Somashekhar M. Punnuri,
Fort Valley State University, United States

REVIEWED BY
Venkatesh Bollina,
Agriculture and Agri-Food Canada (AAFC), Canada
Richard Oteng-Frimpong,
CSIR-Savanna Agricultural Research Institute, Ghana

*CORRESPONDENCE
Dan Makumbi
D.Makumbi@cgiar.org

RECEIVED 04 October 2024
ACCEPTED 20 November 2024
PUBLISHED 09 January 2025

CITATION
Kavai HM, Makumbi D, Nzuve FM, Woyengo VW, Suresh LM, Muiru WM, Gowda M and Prasanna BM (2025) Inheritance of resistance to maize lethal necrosis in tropical maize inbred lines. *Front. Plant Sci.* 15:1506139. doi: 10.3389/fpls.2024.1506139

COPYRIGHT
© 2025 Kavai, Makumbi, Nzuve, Woyengo, Suresh, Muiru, Gowda and Prasanna. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Inheritance of resistance to maize lethal necrosis in tropical maize inbred lines

Hilda M. Kavai [1,2], Dan Makumbi [1*], Felister M. Nzuve [2], Vincent W. Woyengo [3], L. M. Suresh [1], William M. Muiru [2], Manje Gowda [1] and Boddupalli M. Prasanna [1]

[1]International Maize and Wheat Improvement Center (CIMMYT), Nairobi, Kenya, [2]Department of Plant Science and Crop Protection, University of Nairobi, Nairobi, Kenya, [3]Kenya Agricultural and Livestock Research Organization, Non-Ruminant Research Institute, Kakamega, Kenya

Maize (*Zea mays* L.) production in sub-Saharan Africa can be improved by using hybrids with genetic resistance to maize lethal necrosis (MLN). This study aimed to assess the general (GCA) and specific combining ability (SCA), reciprocal effects, and quantitative genetic basis of MLN resistance and agronomic traits in tropical maize inbred lines. A total of 182 hybrids from a 14-parent diallel, along with their parents, were evaluated under artificial MLN inoculation and rainfed conditions for 3 years in Kenya. Disease ratings at four time points, grain yield (GY), and other agronomic traits were analyzed using Griffing's Method 3 and Hayman's diallel models. Significant ($P < 0.001$) GCA and SCA mean squares were observed for all traits under disease conditions and most traits under rainfed conditions, highlighting the importance of both additive and non-additive genetic effects. However, additive gene action predominated for all traits. Narrow-sense heritability estimates for MLN resistance (h^2 = 0.52–0.56) indicated a strong additive genetic component. Reciprocal effects were not significant for MLN resistance, suggesting minimal maternal or cytoplasmic inheritance. Four inbred lines showed significant negative GCA effects for MLN resistance and positive GCA effects for GY under artificial MLN inoculation. Inbred lines CKL181281 and CKL182037 (GCA effects for MLN4 = -0.45 and -0.24, respectively) contained the most recessive alleles for MLN resistance. The minimum number of groups of genes involved in MLN resistance was estimated to be three. Breeding strategies that emphasize GCA could effectively be used to improve MLN resistance in this germplasm.

KEYWORDS

diallel, combining ability, disease resistance, heritability, maize, maternal, reciprocal, maize chlorotic mottle virus

1 Introduction

Maize (*Zea mays* L.) is a major cereal crop in sub-Saharan Africa (SSA), where it covers more than 40 million ha of arable land. It is the most important crop for food security, income, and livelihoods for several million smallholder farmers across SSA, especially in eastern and southern Africa where nearly 85% of the maize produced is used as food (Shiferaw et al., 2011). Maize production in SSA was approximately 70 million metric tons in 2020 (FAOSTAT, 2021) and is largely produced by smallholder farmers. Despite its wide cultivation, the average maize yield in SSA is approximately 2.0 t ha^{-1}, which is far below the global average of approximately 5.8 t ha^{-1} (Erenstein et al., 2022). The low maize yield is attributed to several factors including the frequent occurrence of drought, poor soil fertility, inadequate use of inputs such as improved seed and fertilizers, the impact of pests and diseases (Prasanna et al., 2021), and parasitic weeds (Menkir et al., 2012). Maize diseases of major economic importance in SSA include fungal (Asea et al., 2002; Menkir and Ayodele, 2005; Vivek et al., 2010; Sserumaga et al., 2020) and viral diseases (van Rensburg et al., 1991; Kyetere et al., 1999). Many of the pests and diseases of maize in SSA have become endemic to the region but there have been cases of new transboundary pests and diseases in recent years, affecting the food security and livelihoods of several million resource-constrained smallholder farmers (Prasanna et al., 2020). An example of a transboundary disease occurrence in Africa was the emergence of maize lethal necrosis (MLN) disease in SSA (Wangai et al., 2012; Mahuku et al., 2015a), a disease that was first reported in the Americas in the 1970s (Niblett and Claflin, 1978; Uyemoto et al., 1980).

Maize lethal necrosis was first reported in Kenya in 2011 but has since spread to several other eastern African countries between 2012 and 2018 (Wangai et al., 2012; Mahuku et al., 2015b; Adams et al., 2014; Lukanda et al., 2014; Mudde et al., 2018). This viral disease is caused by the coinfection of maize plants by the *maize chlorotic mottle virus* (MCMV) and any one of the viruses from the family Potyviridae, such as *sugarcane mosaic virus* (SCMV), *maize dwarf mosaic virus* (MDMV), or *wheat streak mosaic virus* (WSMV) (Redinbaugh and Stewart, 2018; Braidwood et al., 2018; Mwatuni et al., 2020). Recent studies revealed that *Johnson grass mosaic virus* (Stewart et al., 2017) and *maize yellow dwarf virus* (Wamaitha et al., 2018) in association with MCMV, cause MLN. The emergence of MLN in eastern Africa is attributed to the entry of MCMV into the region (Prasanna et al., 2020) since the presence of SCMV was reported in maize in East Africa much earlier (Kulkarni, 1973; Louie, 1980).

MLN has seriously affected maize grain yield and production in eastern Africa to varying levels. Based on community surveys in 2013, total maize loss in Kenya due to the MLN outbreak was estimated at 0.5 million metric tons $year^{-1}$ or 22% of the average annual production, with a value of approximately USD 180 million (De Groote et al., 2016). In 2018, the total quantity of maize lost in Kenya was estimated to be approximately 0.17 million metric tons equivalent to approximately USD 51 million (De Groote et al., 2021). Strategies such as crop rotation to break the disease cycle, the use of clean seed, and vector control have been proposed to manage MLN in SSA. The most economically viable and environmentally sustainable approach to control and manage MLN is the development of resistant or tolerant maize varieties. The economic value of adopting MLN tolerant hybrids was estimated at USD 195–678 million in Kenya and USD 245–756 million in Ethiopia depending on adoption levels of 25–75% (Marenya et al., 2018), suggesting a considerable benefit to farmers in utilizing MLN resistant varieties.

The International Maize and Wheat Improvement Center (CIMMYT) in collaboration with national partners, initiated screening of its germplasm stock and from other sources for resistance to MLN in 2012. A few sources of MLN resistance were identified, and introgression of resistance into CIMMYT's elite germplasm was initiated. The key lines that have been used for introgression of MLN resistance alleles into CIMMYT's mid-altitude adapted maize germplasm are the yellow lines KS23-5 and KS23-6 from Kasetsart University, Thailand (Jones et al., 2018; Prasanna et al., 2020). These two lines were extracted from KS23(S)C5, a population that had undergone five cycles of S1 recurrent selection (Jampatong et al., 2010) and have resistance to maize mosaic virus (Brewbaker, 2009). Through pedigree breeding and the use of doubled haploid (DH) technology, several inbred lines have been developed, screened under MLN disease pressure to identify resistant lines with adaptive traits, and utilized for hybrid development. Some of the elite CIMMYT lines have also been converted into MLN resistant versions using the major quantitative trait locus (QTL) (qMLN06.157) from KS23-6 (Prasanna et al., 2020; Murithi et al., 2021).

Knowledge of the genetic basis of resistance to diseases is important in developing breeding strategies. Diallel studies have been used to investigate the genetics of virus disease resistance in maize (Josephson and Naidu, 1971; Loesch and Zuber, 1972; Naidu and Josephson, 1976; Rosenkranz and Scott, 1987; Mutengwa et al., 2012; Beyene et al., 2017; Nyaga et al., 2020). Beyene et al. (2017) and Nyaga et al. (2020) reported that additive gene action is more important than nonadditive gene action for MLN resistance and identified some MLN tolerant inbred lines. Studies on the genetics of maize virus diseases have focused mainly on combining ability for disease parameters but not reciprocal effects that could impact hybrid development plans and disease resistance improvement strategies. To date, studies on MLN in tropical maize have used early to intermediate maturity germplasm. The upper mid-altitude ecologies of Eastern Africa are a major maize production area, where late-maturity maize is the preferred type. However, MLN disease is increasingly affecting this region. Therefore, incorporating resistance to MLN into late-maturity maize germplasm is crucial for sustaining production in this region. Using two different sources of resistance to MLN, alleles for resistance have been introgressed into CIMMYT's late-maturity maize germplasm suitable for the upper

Abbreviations: CIMMYT, Centro Internacional de Mejoramiento de Maíz y Trigo (International Maize and Wheat Improvement Center); DTA, days to anthesis; DTS, days to silking; EPP, ears per plant; ESA, eastern and southern Africa; GCA, general combining ability; GY, grain yield; MCMV, maize chlorotic mottle virus; MLN, maize lethal necrosis; SCA, specific combining ability; SCMV, sugarcane mosaic virus; SSA, sub-Saharan Africa; TLB, *Turcicum* leaf blight.

10.3389/fpls.2024.1506139

mid-altitude ecology of Eastern Africa. Limited to no information on the combining ability and quantitative genetic parameters of MLN resistance in adapted late maturing tropical maize germplasm has been reported. The objectives of this study were to (i) estimate the combining ability of MLN resistance and other traits among 14 late-maturity inbred lines and assess the importance of reciprocal effects, and (ii) investigate the quantitative genetic basis of MLN resistance in tropical maize.

2 Materials and methods

2.1 Genetic material

Fourteen inbred lines with varying response to MLN and other characteristics based on field evaluations were selected for this study (Table 1). The selected genotypes included five lines conventionally developed through pedigree breeding from biparental crosses involving a known source of resistance to MLN (entries 1–5), six DH lines (entries 6–11), and three drought tolerant inbred lines (entries 12–14). The 14 inbred lines were crossed in a full diallel mating design with reciprocals to generate 182 F_1 hybrids. The crossing was performed at the Kenya Agricultural and Livestock Research Organization (KALRO) Kiboko Research Center (2°15'S, 37°75'E, 975 m asl), Kenya, in 2019. In the same year, seed of the 14 parental lines and one MLN resistant line (KS23-6) was increased to compose the line trials.

2.2 Test locations, experimental design, and trial management

The 182 F_1 hybrids plus four commercial check hybrids were grown in seven trials that were planted at two locations in Kenya in 2020, 2021, and 2022. The hybrid trial was laid out as a 3 × 62 alpha-lattice (Patterson and Williams, 1976) with two replications. A line evaluation trial was also formed, consisting of the 14 parental lines of the diallel hybrids and one inbred line check. The inbred line trial was laid out as a 3 × 5 alpha-lattice with two replications. In both trials, each experimental unit consisted of one row 5 m long, spaced 0.75 m apart and 0.25 m between plants, resulting in a population density of approximately 53,333 plants ha^{-1}. The hybrid and parental line trials were evaluated at the KALRO-CIMMYT MLN screening facility at Naivasha (0°43'S, 36°26'E, 2086 m asl) under artificial inoculation with MLN. There were eight trials (four each of hybrids and lines) planted at Naivasha. The same set of germplasm was evaluated at the KALRO Kakamega Non-Ruminant Research Center (0°16'N, 34°49'E, 1585 m asl) under rainfed and natural

TABLE 1 List of 14 inbred lines used to develop F_1 hybrids and their reciprocal crosses in a full diallel.

No.	Name	Pedigree	Origin	Growing degree-days (GDD)	Characteristics
1	CKL18912	(CKL05003/KS23-6)-B-9-4-5-2-2-B-B-B	Kenya	970	Tolerant MLN
2	CKL181281	(((KU1403 x 1368)-7-2-1-1-B-B/CML444)-B-8-7-3-2-6-1-2-B-B-B/KS23-6)-B-1-1-4-1-2-B-B-B	Kenya	812	Resistant to MLN
3	CKL181379	(((KU1403 x 1368)-7-2-1-1-B-B/CML444)-B-8-7-3-2-6-1-2-B-B-B/KS23-6)-B-22-3-1-2-3-B-B-B	Kenya	829	Resistant to MLN
4	CKL181847	((CKL05003/La Posta Seq C7-F64-2-6-2-2-B-B-B)DH110-B-B/KS23-6)-B-17-4-1-1-1-B-B-B	Kenya	874	Tolerant to MLN
5	CKL182037	(([CML444/CML395//DTPWC8F31-1-1-2-2-BB]-4-2-2-1-1-B*4/(9071xBabamgoyo)-3-1-BBB)-B-1-2-3-1-3-B/KS23-6)-B-2-1-3-2-1-B-B-B	Kenya	851	Tolerant to MLN
6	CKL176616	((([LZ956441/LZ966205]-B-3-4-4-B-5-B*7/LaPostaSeqC7-F71-1-2-1-1-BBB)-1-7-1-1-BB-B/KS23-5)-B)DH15-B-B-B-B-B	Kenya	926	Tolerant to MLN
7	CKL175951	(CML495/CML341)DH23-B-B-B-B-B	Mexico	903	Late, lowland, drought tolerant (DT)
8	CKL175755	(CML341/CML247)DH84-B-B-B-B-B	Mexico	926	Late, lowland, DT
9	CKL175798	(CML343/CML495)DH57-B-B-B-B-B	Mexico	829	Late, lowland, DT
10	CKL176082	(CML495/CML341)DH3-B-B-B-B-B	Mexico	887	Late, lowland
11	CKDHL120918	(CML445/ZM621B//P100C6-200-1-1-B-B-B-B)@3020-B-B-B-B	Kenya	846	Tolerant to MLN
12	CML585	[KILIMA(ST94)-S5:115/[M37W/ZM607#BF37SR …]]-B-B-3-5-B-B*4	Kenya	903	Susceptible to MLN; DT
13	CKL14546	(CKL05017/LaPostaSeqC7-F78-2-1-1-1-B-B-B)-B-2-1-2-1-1-B-B-B-B	Kenya	893	Susceptible to MLN; DT
14	CML444	P43-C9-1-1-1-1-1-B-B-B	Zimbabwe	904	Susceptible to MLN; DT

foliar disease pressure conditions in six trials (three each for hybrids and lines). The inbred line trials were planted side by side with the hybrid trial at both locations. Standard agronomic and cultural practices were performed as recommended for each location.

2.3 Artificial inoculation with MLN causing viruses (MCMV and SCMV) and disease rating

The pure mother cultures of MCMV and SCMV were maintained on susceptible host maize hybrids H614 and PHB30G19, respectively, in separate insect-proof net houses at the KALRO-CIMMYT MLN screening facility. The inoculum was prepared following the protocol described in detail in previous studies (Gowda et al., 2018; Sitonik et al., 2019). Briefly, SCMV and MCMV inocula were initially prepared separately. Then, at the time of inoculation, the two viruses were mixed at a ratio of 4:1 of SCMV and MCMV, respectively. The hybrid and inbred line trials were inoculated with the mixture of SCMV and MCMV twice: first at the 4–5 leaf stage, and a second inoculation was carried out seven days after the first inoculation. A motorized backpack mist blower (Solo 423 Mist Blower, 12 L capacity) was used to deliver the inoculum at a pressure of 10 kg cm^{-2}.

Disease rating for response to MLN infection was visually done by observing disease symptoms on all plants in a plot at four time points: at 21 (MLN1), 28 (MLN2), 35 (MLN3), and 42 (MLN4) days after the first inoculation for both inbred lines and hybrids. A scale of 1–9 was used for disease rating, where 1 = completely clean plants with no visible MLN disease symptoms, 3 = mild chlorotic streaks on emerging leaves, 5 = chlorotic streaks and mottling throughout the plant, 7 = severe chlorotic mottling, mosaic, and leaf necrosis throughout the plant, and 9 = complete plant necrosis, and dead plants (Prasanna, 2021; https://hdl.handle.net/10883/21703). The four MLN disease ratings were used to calculate the area under disease progress curve (AUDPC) which is a quantitative measure of disease intensity with time as follows:

$$\text{AUDPC} = \sum_{i=1}^{n}\left[\frac{Y_i + Y_{i+1}}{2}\right](T_{i+1} + T_i)$$

where i = time of MLN disease rating, T_i is the number of days after inoculation, and Y_i is the MLN disease rating (Shaner and Finney, 1977).

2.4 Agronomic and foliar disease data

Days to anthesis (DTA, recorded as days from planting to when 50% of the plants started to shed pollen), days to silking (DTS, recorded as days from planting to when 50% of the plants had emerged silks) and ears per plant (EPP) were recorded. The number of ears per plant (EPP) was obtained by dividing the total number of ears per plot by the number of plants harvested. The response to Turcicum leaf blight (TLB) a major foliar disease in SSA caused by *Exserohilum turcicum* (Pass.) Leonard & Suggs in SSA was recorded under heavy natural disease pressure at KALRO-Kakamega on a scale of 1–9, where 1 = highly resistant, no disease symptoms, and 9 = highly susceptible, with severely necrotic leaves. Kakamega is a high natural disease pressure location used for assessing the response to major maize foliar diseases (Vivek et al., 2010). The foliar disease response was recorded when the crop was at the dough stage. All ears in a single-row plot were harvested, weighed, and representative samples of ears were shelled to determine the percent moisture using a Dickey-John multigrain moisture tester (DICKEY-John Corporation, IL, USA). The grain yield, expressed as t ha^{-1} was calculated from cob weight assuming a shelling percentage of 80% and adjusted to 12.5% moisture content.

2.5 Statistical analyses

2.5.1 Analysis of variance

The data were first assessed for homogeneity of variance using Levene's test before ANOVA, and variances were found to be homogeneous. Analyses of variance were performed using META-R (Alvarado et al., 2020), first by location and then across each separate management condition (artificial MLN inoculation and rainfed conditions). Each location-year combination was considered an environment. Genotypes and locations were considered fixed and random effects, respectively. The linear model used for combined analysis across environments was as follows:

$$Y_{ijrk} = \mu + \alpha_i + \beta_j + \rho_r(\beta_j) + \lambda_k[\rho_r(\beta_j)] + \alpha\beta_{ij} + \varepsilon_{ijrk}$$

where Y_{ijrk} is the mean of the ith genotype, in the rth replicate within the kth sub-block of the jth environment; μ is the grand mean; α_i is the effect of the ith genotype; β_j is the effect of the jth environment; $\rho_r$is the effect of the rth replicate; $\rho_r(\beta_j)$ is the effect of the replicates within environments; λ_k is the effect of the kth incomplete block; $\lambda_k[\rho_r(\beta_j)]$ is the effect of the incomplete blocks within replicates and environments; $\alpha\beta_{ij}$ is the effect of the genotype × environment interaction; and ε_{ijrk} is the residual error. To estimate variance components, all factors were considered random effects. The best linear unbiased predictions (BLUPs) and best linear unbiased estimates (BLUEs) for the genotypes were computed. The broad-sense heritability of recorded traits and disease parameters across environments was estimated according to Hallauer et al. (2010) as follows:

$$H^2 = \frac{\sigma_G^2}{\sigma_G^2 + \frac{\sigma_{GE}^2}{e} + \frac{\sigma_e^2}{er}}$$

in which σ_G^2, $\sigma_{G\times E}^2$, and σ_e^2 are the genotype, genotype × environment, and residual variance components, respectively, E is the number of environments, and r is the number of replications.

2.5.2 Diallel analysis

Data from the hybrid trial, excluding that of the commercial hybrid checks, were subjected to diallel analysis following Griffing's Method 3 Model 1 (Griffing, 1956). The use of Method 3 of Griffing allowed us to investigate the possible influence of reciprocal effects

10.3389/fpls.2024.1506139

due to cytoplasmic differences and/or cytoplasmic-genic relationships on MLN parameters and other traits. The hybrid source of variation was partitioned into general (GCA) and specific combining ability (SCA), and reciprocal effects. The reciprocal effects were further partitioned into maternal and nonmaternal effects. Diallel analysis was carried out using the AGD-R software for R v3.0 (Rodríguez et al., 2020). The following linear model was used for the analysis:

$$X_{ijkt} = \mu + t_i + b_{ki} + v_{ij} + (tv)_{ijt} + e_{ijkt}$$

where X_{ijkt} = observed trait value (i and j, are parents; k, replication; t, environment), μ = population mean; t_i = environment effect; b_{ki} = block or replication within environment effect; v_{ij} = genotype effect = $g_i + g_j + s_{ij} + r_{ij}$ [where g_i = GCA effect of the ith parent, g_j = GCA effect of the jth parent, s_{ij} = SCA effect of the ijth F_1 hybrid, r_{ij} = reciprocal effect of the ijth or jith F_1 hybrid = $m_i + m_j + n_{ij}$(where m_i = maternal effect of parental line i, m_j = maternal effect of parental line j, and n_{ij} = nonmaternal effect of the ijth or jith F_1 hybrid], $(tv)_{ijt}$ = interaction between genotypes and environments, e_{ijkt} = residual effect. The relative importance of GCA and SCA was assessed using the ratio of the GCA and SCA sums of squares.

The F_1 hybrid and parental inbred line data (excluding that of the inbred line check) for MLN disease resistance parameters were further subjected to Hayman's model (Hayman, 1954a, b) of diallel analysis. Hayman's diallel analysis provides genetic information on additive and dominance effects of genes, average degree of dominance, distribution of genes, and number of groups of genes which control a trait among others. Hayman's diallel analysis involves graphical and statistical analyses of array variances and covariances and estimation of genetic parameters (Hayman, 1954a, b; Mather and Jinks, 1971). Briefly, Hayman's diallel analysis requires the calculation of the variances from all crosses of each parental array (V_r), and the covariance between parents and their crosses in each array (W_r) (Hayman, 1954a). The variances and covariances were calculated and used to construct a W_r-V_r graph. In addition, quantitative genetic parameters for MLN resistance were estimated. Hayman's diallel analysis was carried out using the SASHAYDIALL program (Makumbi et al., 2018a) in SAS (SAS Institute, 2016).

3 Results

3.1 ANOVA under artificial MLN conditions

The combined ANOVA across four seasons under artificial MLN inoculation showed significant ($P < 0.001$) environment (E) and genotype (G) mean squares for all agronomic traits and disease parameters (Table 2). The G × E interaction was significant for all traits except EPP. Both GCA and SCA mean squares were significant ($P < 0.001$) for all the traits measured under artificial

TABLE 2 Mean squares from combined analysis (Griffing's Method 3 Model 1) for agronomic traits and MLN disease resistance parameters in a 14-parent diallel evaluated under artificial MLN inoculation at Naivasha over 3 years (2020–2022).

Source of variation	df	DTA	EPP	GY	MLN1	MLN2	MLN3	MLN4	AUDPC
Environment (E)	3	5643.69***	7.24***	246.55***	19.22***	16.30***	27.94***	49.26***	9626.44***
Rep(E)	4	12.05***	1.16***	0.51	0.04	0.42***	0.58***	0.45***	132.19***
Genotypes (G)	181	47.27***	0.84***	10.05***	3.47***	5.28***	6.32***	6.71***	2351.17***
GCA	13	351.74***	5.87***	64.02***	25.36***	39.33***	50.05***	53.22***	18188.86***
SCA	77	20.94***	0.98***	11.83***	3.73***	5.61***	6.26***	6.71***	2427.19***
Reciprocal	91	11.39***	0.22	0.97***	0.17	0.19	0.23	0.22	59.87*
Maternal (M)	13	11.80***	0.17	0.83	0.13	0.20	0.18	0.18	58.26
Nonmaternal (NM)	78	11.32***	0.23	1.00***	0.18	0.19	0.24	0.22	60.13*
G × E	542	7.02***	0.29	0.92***	0.28***	0.29***	0.38***	0.40***	104.57***
GCA × E	39	3218.49***	0.77***	4.05***	0.69***	0.72***	0.46***	0.76***	190.28***
SCA × E	231	807.56***	0.31	0.89***	0.22**	0.17	0.20	0.22	41.89
Reciprocal × E	273	4.86**	0.25	0.64	0.18	0.16	0.22*	0.22	54.41*
M × E	39	433.99***	0.19	0.66	0.29**	0.34***	0.52***	0.60***	157.90***
NM × E	234	0.00	0.26	0.64	0.17	0.13	0.18	0.15	37.16
Error	230	3.41	0.26	0.59	0.15	0.15	0.18	0.18	43.11
GCA: SCA ratio		0.85	0.68	0.65	0.70	0.70	0.73	0.73	0.72

*, **, and *** indicate significance at the 0.05, 0.01, and 0.001 levels, respectively.
DTA, days to anthesis; EPP, ears per plant; GY, grain yield; MLN1, MLN2, MLN3, and MLN4, maize lethal necrosis disease rating at 21, 28, 35 and 42 days after inoculation, respectively; AUDPC, area under disease progress curve.

MLN inoculation. The differences between the F_1 hybrids and their reciprocals were significant for DTA, GY, and AUDPC. Partitioning of the reciprocal source of variation into maternal and nonmaternal effects revealed that maternal effects were significant ($P < 0.001$) for only DTA, while nonmaternal effects were significant for DTA, GY, and AUDPC (Table 2). Furthermore, the GCA × E interaction was significant for both agronomic and disease parameters, while the SCA × E interaction was significant for only DTA, GY, and MLN1. The reciprocal × E interaction was significant for DTA, MLN3 and AUDPC, while the maternal × E interaction was significant for DTA, MLN disease resistance parameters, and AUDPC. The GCA: SCA ratio varied for the agronomic traits and ranged from 0.65 to 0.85 for GY and agronomic traits, and from 0.70 to 0.73 for the MLN disease resistance parameters.

3.2 ANOVA under rainfed conditions

The combined ANOVA under rainfed conditions revealed significant ($P < 0.001$) environment (E), genotype (G), and G × E interaction mean squares for all traits except the EPP genotype mean squares (Table 3). The results revealed that both GCA and SCA mean squares were significant ($P < 0.001$) for all the traits except EPP, which showed only significant GCA mean squares. The reciprocal differences between the F_1 hybrids and their crosses were significant for DTA, EPP, and GY. Both maternal and nonmaternal effects were significant for DTA, and significant maternal effects were recorded for TLB. Nonmaternal effects were significant for GY and EPP. The GCA × E interaction was significant for all traits, while SCA × E interaction was significant for EPP and GY. The reciprocal × E and maternal × E interactions were significant for flowering traits DTA and DTS. The GCA: SCA ratio ranged from 0.50 to 0.88 for the agronomic traits and was 0.94 for TLB.

3.3 Performance of hybrids

The mean GY of the F_1 hybrids under artificial MLN inoculation was 2.0 t ha^{-1}, with a range of 0.4 to 6.4 t ha^{-1}, while the GY of the commercial check hybrids ranged from 0.9 to 3.4 t ha^{-1} (Table 4). The range of the second disease rating taken 28 days after inoculation (MLN2) was 2.5 to 6.3, while the fourth disease rating taken 42 days after inoculation (MLN4) was 3.4 to 7.7. The variance due to genetic effects was 2.5 times greater than the residual variance for GY. For the MLN disease scores, the variance due to genetic effects was 3.9 to 4.2 times greater than the residual variance. Broad-sense heritability estimates were high for most traits (0.89–0.95) and moderate for EPP (0.67). The hybrid with the highest yield under artificial MLN inoculation was the reciprocal cross P6 × P2 (6.4 t ha^{-1}), which also had the lowest disease scores for MLN3 (2.8) and MLN4 (3.4), and lowest AUDPC (55.4) (Supplementary Table S1). Grain yield was negatively correlated with all four MLN disease ratings (r = -0.80 to -0.86, $P < 0.001$). Under rainfed conditions, the mean GY was 4.0 t ha^{-1} and 4.8 t ha^{-1} for the F_1 hybrids and commercial check hybrids, respectively (Table 4). Overall, the GY ranged from 2.5 to 6.0 t ha^{-1} under rainfed conditions. The mean number of DTA was greater at

TABLE 3 Mean squares from combined analysis (Griffing's Method 3 Model 1) for agronomic traits and Turcicum leaf blight in a 14-parent diallel evaluated under rainfed conditions at Kakamega, 2020–2022.

Source of variation	df	DTA	DTS	EPP	GY	TLB
Environment (E)	2	1684.04***	2170.82***	2.02**	1699.85***	1158.48***
Rep(E)	3	25.26***	34.34***	0.10	11.80***	0.77***
Genotypes (G)	181	36.08***	40.49***	0.06	6.86***	5.10***
GCA	13	369.37***	372.89***	0.14***	34.47***	62.96***
SCA	77	17.21***	25.28***	0.05	6.89***	1.38***
Reciprocal	91	5.51***	7.08	0.06*	3.04*	0.74
Maternal (M)	13	8.05***	9.14	0.06	2.99	1.36**
Nonmaternal (NM)	78	5.08***	6.74	0.06*	3.04*	0.63
G × E	357	5.91***	8.29***	0.06*	3.29**	0.95**
GCA × E	26	296.39***	295.15***	0.14***	5.96***	6.98***
SCA × E	154	0.0	0.0	0.07***	3.13**	0.53
Reciprocal × E	182	4.88***	7.78**	0.05	2.69	0.59
M × E	26	142.79***	155.74***	0.07*	2.70	0.84
NM × E	156	0.0	0.0	0.04	2.69	0.55
Error	168	2.65	5.43	0.04	2.35	0.58
GCA: SCA ratio		0.88	0.83	0.50	0.65	0.94

*, **, and *** indicate significance at the 0.05, 0.01, and 0.001 levels, respectively.
AUDPC, Area under disease progress curve; DTA, days to anthesis; DTS, days to silking; EPP, ears per plant; GY, grain yield; TLB, Turcicum leaf blight.

10.3389/fpls.2024.1506139

TABLE 4 Summary statistics, variance component and heritability estimates for agronomic traits, MLN disease resistance parameters, and area under disease progress curve (AUDPC) of 182 maize hybrids and four commercial hybrid checks evaluated under artificial inoculation with MLN at Naivasha, and under rainfed conditions at Kakamega for three seasons, 2020–2022.

Trait	Unit	Mean	Range	$LSD_{0.05}$	Genotypic variance	G × E variance	Residual variance	Heritability
Naivasha (artificial MLN inoculation)								
GY (F_1 hybrids)	t ha^{-1}	2.0	0.4–6.4	1.0	1.44	0.10	0.77	0.92
GY (Checks)		2.2	0.9–3.4					
DTA	days	92	84–97	2.5	6.38	1.34	3.38	0.89
EPP	no.	0.9	0.4–1.9	0.4	0.07	0.01	0.26	0.67
MLN1	1–9	3.5	2.1–4.9	0.5	0.40	0.05	0.16	0.92
MLN2	1–9	4.4	2.5–6.3	0.5	0.63	0.06	0.16	0.95
MLN3	1–9	5.0	2.8–7.3	0.6	0.75	0.08	0.18	0.94
MLN4 (F1 hybrids)	1–9	5.4	3.4–7.7	0.6	0.80	0.10	0.19	0.94
MLN4 (Checks)		5.8	4.8–6.8					
AUDPC (F_1 hybrids)		193.9	110.7–279.2	19.9	1133.51	111.50	180.99	0.96
AUDPC (Checks)		202.7	173.7–245.2					
Kakamega (rainfed)								
GY (F_1 hybrids)	t ha^{-1}	4.0	2.5–6.0	1.5	0.60	0.41	2.07	0.56
GY (Checks)		4.8	4.1–5.4					
DTA	days	80	75–86	2.5	5.45	0.93	2.97	0.87
DTS	days	80	75–85	3.0	5.77	0.82	5.65	0.83
EPP	no.	1.1	0.8–1.3	-	0.00	0.01	0.04	0.04
TLB (F_1 hybrids)	1–9	4.1	2.4–5.8	1.0	0.67	0.17	0.53	0.82
TLB (Checks)		4.1	2.7–4.6					

AUDPC, area under disease progress curve; DTA, days to anthesis; DTS, days to silking; EPP, ears per plant; GY, grain yield; MLN1, MLN2, MLN3 and MLN4, maize lethal necrosis disease rating at 21, 28, 35 and 42 days after inoculation, respectively; TLB, Turcicum leaf blight.

Naivasha (2086 m asl) than at Kakamega (1585 m asl), possibly due to the higher elevation and cooler environment at Naivasha. The TLB disease rating ranged from 2.4 to 5.8. Broad-sense heritability was moderate for GY (0.56) and high for agronomic traits and TLB (0.82–0.86), except for EPP. The top hybrid for GY under rainfed conditions at Kakamega was P12 × P2 (6.0 t ha^{-1}), with a TLB score of 2.9 (Supplementary Table S2).

3.4 ANOVA for inbred lines and *per se* performance

Combined ANOVA revealed significant ($P < 0.001$) genotype (G), and G × E interaction mean squares for MLN disease ratings and AUDPC under artificial MLN inoculation while significant genotype mean squares were recorded for TLB under rainfed conditions (Supplementary Table S3). The MLN disease rating ranged from 2.1 to 6.0, 2.1 to 7.4, 2.1 to 8.5, and 2.1 to 8.6 for MLN1, MLN2, MLN3 and MLN4, respectively. The most MLN resistant line with a score of 2.1 was CKL181281 (Table 5; Figure 1). The TLB disease rating for these lines ranged from 2.4 to 4.8. Broad-sense heritability was high for the MLN disease ratings, AUDPC and TLB (0.87–0.96) and moderate for DTA.

3.5 Estimates of GCA and SCA effects

The GCA effects varied between parents for the different traits. Inbred lines CKL181281, CKL181379, CKL182037, CKL176616, and CKL175755 had significant negative GCA effects for all four MLN disease resistance parameters and AUDPC, and therefore contributed to MLN resistance in their hybrids (Table 5). Inbred line CKL176616 had the largest significant negative GCA effect for both MLN4 (-0.81, $P < 0.001$) and AUDPC (-14.54, $P < 0.001$), followed by CKL175755 for MLN4 (-0.63, $P < 0.001$). Another inbred line, CKL181847, had significant negative GCA effects for two of the four MLN disease ratings (MLN1 and MLN2) and AUDPC. Three inbred lines (CKL175951, CKL175798, and CKL176082) had significant negative GCA effects for the fourth MLN disease rating. In contrast, five inbred lines showed significant positive GCA effects for three MLN ratings (MLN2–MLN3) and AUDPC, suggesting that these lines contributed to MLN

10.3389/fpls.2024.1506139

TABLE 5 General combining ability (GCA) estimates for maize lethal necrosis disease ratings, AUDPC, and agronomic traits, and *per se* performance of 14 maize inbred lines under artificial MLN inoculation at Naivasha over 3 years (2020–2022).

Line	Name	MLN1	MLN2	MLN3	MLN4	AUDPC	DTA (days)	EPP (No)	GY (t ha^{-1})	MLN3	MLN4	AUDPC
		(1–9)								(1–9)		
1	CKL18912	0.27***	0.28***	0.45***	0.56***	8.01***	1.88***	-0.16**	-0.24***	4.8	6.0	93.4
2	CKL181281	-0.49***	-0.57***	-0.53***	-0.45***	-11.05***	-2.34***	0.17***	0.80***	2.1	2.1	43.2
3	CKL181379	-0.29***	-0.44***	-0.39***	-0.35***	-8.06***	-0.58***	0.04	0.09	3.4	3.7	65.9
4	CKL181847	-0.13***	-0.19***	-0.04	0.07	-1.82***	1.09***	0.04	0.01	3.7	4.4	72.1
5	CKL182037	-0.25***	-0.35***	-0.36***	-0.24***	-6.66***	-1.77***	0.07	0.69***	3.4	3.7	65.4
6	CKL176616	-0.51***	-0.62***	-0.80***	-0.81***	-14.54***	0.42**	0.25***	1.23***	2.9	3.3	58.6
7	CKL175951	0.01	-0.01	-0.02	-0.16***	-0.74	0.65***	0.01	-0.29***	5.2	5.7	103.6
8	CKL175755	-0.13***	-0.23***	-0.55***	-0.63***	-8.14***	0.73***	0.06	-0.08	4.7	5.0	95.0
9	CKL175798	-0.07	-0.05	-0.03	-0.19***	-1.45*	-0.21	0.07	0.26***	5.1	5.4	98.8
10	CKL176082	-0.05	-0.01	-0.02	-0.13**	-0.86	1.44***	0.02	-0.33***	5.5	6.2	105.0
11	CKDHL120918	0.04	0.24***	0.28***	0.29***	4.74***	-1.89***	0.09*	0.02	4.3	5.0	79.9
12	CML585	0.81***	0.93***	1.07***	1.11***	20.69***	-0.88***	-0.48***	-1.02***	8.5	8.6	163.3
13	CKL14546	0.31***	0.47***	0.41***	0.41***	8.65***	-0.15	-0.05	-0.37***	6.0	6.6	111.0
14	CML444	0.47***	0.58***	0.53***	0.52***	11.22***	1.59***	-0.12***	-0.66***	6.7	6.7	123.4
15	***KS23-6***[†]									***3.9***	***4.2***	***74.3***
SE/LSD$_{(0.05)}$[$]		0.03	0.03	0.03	0.03	0.46	0.13	0.04	0.05	0.8	1.0	13.5

*, **, and *** indicate significance at the 0.05, 0.01, and 0.001 levels, respectively.
[†]KS23-6, an MLN resistant donor line was used as a check in the line trial.
[$]SE of GCA effects; LSD for per se performance.
AUDPC, area under disease progress curve; DTA, days to anthesis; EPP, ears per plant; GY, grain yield; MLN1, MLN2, MLN3 and MLN4, maize lethal necrosis disease rating at 21, 28, 35 and 42 days after inoculation, respectively.

susceptibility in their hybrids. Inbred lines CKL181281, CKL182037, CKL176616, and CKL175798 exhibited significant positive GCA effects for GY under artificial MLN inoculation. We computed the nonparametric Spearman's rank correlation coefficient between the GCA effects for the four MLN disease resistance parameters, and the results revealed very strong correlations among the scores (r_s = 0.93–0.98, $P < 0.001$). In terms of SCA effects, the results showed that nearly a similar number of hybrids had significant negative SCA effects on the second (35), third (37), and fourth (35) MLN disease rating (Supplementary Tables S4, S5). Several hybrid combinations, such as CKL181281 × CKL182037, CKL18912 × CKL181281, CKL18912 × CKL182037, and CKL181379 × CKL182037 consistently exhibited negative significant SCA effects for MLN2, MLN3, and MLN4, suggesting their potential for MLN disease resistance. For GY, 31 hybrids (33%) exhibited significant positive SCA effects under artificial MLN inoculation. Twenty-nine hybrids had significant negative SCA effects for the fourth MLN disease rating and significant positive SCA effects for GY. The top-yielding hybrids were between parents with desirable GCA effects for MLN resistance and GY. For example, the top two hybrids (P6 [CKL176616] × P2 [CKL181281], SCA = 2.0 t ha^{-1} and P6 [CKL176616] × P5 [CKL182037], SCA = 1.2 t ha^{-1}) were between lines with significant GCA effects for both traits. Furthermore, P6 (CKL176616), P5 (CKL182037) and P2 (CKL181281) were the parents of 18 of the top 25 hybrids in terms of GY performance under MLN inoculation.

Under rainfed conditions, the inbred lines CKL18912, CKL175951, CKL175755, CKL176082, and CML444 had significant positive GCA effects for both maturity parameters (Table 6). Five inbred lines (CKL181281, CKL182037, CKDHL120918, CML585, and CKL14546) exhibited significant positive GCA effects for GY under rainfed conditions. Six inbred lines showed significant desirable GCA effects for TLB. Inbred lines CKL181281 and CKL182037, which showed significant and desirable GCA effects for MLN resistance and GY under both artificial MLN inoculation and rainfed conditions, also exhibited significant negative GCA effects for TLB. The results indicated that 10 hybrids had significant positive SCA effects for GY, while 10 hybrids had significant negative SCA effects for TLB (Supplementary Table S6). Under rainfed conditions, either P2 (CKL181281) or P5 (CKL182037) was the parent of 11 of the top 25 hybrids for GY.

3.6 Genetic parameters of MLN resistance

The quantitative genetic parameters for MLN disease resistance were studied using Hayman (1954a) diallel analysis model. The ANOVA revealed that both additive (a) and dominant (b) gene

10.3389/fpls.2024.1506139



FIGURE 1
Inbred line CKL181281 under artificial MLN inoculation at the KALRO-CIMMYT MLN Screening Facility at Naivasha. This line was the most resistant in this study and was rated 2.1 on a scale of 1-9. On the right is an MLN susceptible inbred line.

effects were significant in the control of MLN resistance, but additive gene effects were more important in the inheritance of MLN resistance (Table 7). Furthermore, the analysis revealed that neither maternal (c) nor reciprocal effects (d) were significant for the four MLN disease resistance parameters, indicating that there were no significant differences between reciprocal crosses. This result is consistent with Griffing's diallel analysis for the same traits (Table 2). The results show that there was significant directional dominance (b_1). A comparison of the hybrid and parental means shows that dominance was for susceptibility to MLN. The interactions a × E, b × E, c × E, and d × E were significant for the four MLN disease resistance parameters, suggesting that gene effects were influenced by the environment. The genetic parameter estimates for the four MLN disease ratings are presented in Table 8. The additive and dominance variance components were both significant for the four MLN disease resistance parameters, but the additive variance component was of greater magnitude. This suggested a greater role of additive gene action in MLN resistance. Gene frequency asymmetry was detected but was less important for all four disease parameters ($H_2/4H_1 <$ 0.20). The mean degree of dominance was 0.89 for MLN1, 0.93 for MLN2, 0.85 for MLN3, and 0.89 for MLN4, indicating incomplete dominance for MLN resistance. The minimum number of groups of genes for MLN resistance was estimated to be 3.53, 3.82, 3.63, and 3.31 for MLN1, MLN2, MLN3, and MLN4, respectively. The correlation coefficient between $W_r + V_r$ and Y_r (parental mean) was negative (−0.64 to −0.59) for the MLN parameters, suggesting that the dominant genes increased susceptibility to MLN. Narrow-sense heritability estimates ranged from 0.52 to 0.56 for the four MLN disease resistance parameters.

The variance (V_r) and covariance (W_r) estimates for the third and fourth MLN disease rating (when full expression of the response of lines to MLN was best observed) were used for regression analysis. The slope of the regression line was >1.0, suggesting adequacy of the model (Figure 2). The lines were spread along the regression line, which suggested diversity among the lines for MLN resistance. Based on the W_r-V_r plot, the inbred lines used in this study could be grouped into four groups. Group 1 included lines 7 (CKL175951), 8 (CKL175755), 9 (CKL175798), and 10 (CKL176082) which had more dominant alleles. Group 2 included lines 11 (CKDHL120918), 13 (CKL14546), and 14 (CML444), which had slightly more recessive alleles than did the lines in group 1. Group 3 was composed of lines 1 (CKL18912), 3 (CKL181379), 4 (CKL181847), and 6 (CKL176616), which had more recessive than dominant alleles (75:25). This group of lines showed negative and significant GCA effects for MLN disease resistance parameters and AUDPC, except for line 1 (CKL18912). Inbred lines 2 (CKL181281) and 5 (CKL182037) formed the fourth group and contained the most recessive alleles for MLN resistance.

4 Discussion

To effectively combat the spread of MLN in SSA, breeding for resistant varieties coupled with other strategies like clean seed as part of an integrated disease management strategy is important. The

10.3389/fpls.2024.1506139

TABLE 6 General combining ability (GCA) estimates for agronomic traits, grain yield, and TLB, and *per se* performance of 14 maize inbred lines under rainfed conditions in Kakamega over 3 years (2020–2022).

Line	Name	GCA Effects					*Per se*
		DTA (days)	DTS (days)	EPP (No)	GY (t ha^{-1})	TLB (1-9)	TLB (1-9)
1	CKL18912	1.35***	1.89***	-0.02	-0.44**	0.21**	3.4
2	CKL181281	-1.14***	-0.37	0.03	0.62***	-0.87***	2.4
3	CKL181379	-0.38**	-0.49*	-0.01	-0.66***	-0.93***	2.4
4	CKL181847	0.24	1.20***	0.01	-0.23	0.17**	2.9
5	CKL182037	-0.60***	-0.23	-0.03	0.29*	-0.51***	2.7
6	CKL176616	-0.20	-0.32	-0.04*	-0.78***	1.26***	4.8
7	CKL175951	1.19***	0.45*	0.02	0.02	-0.01	2.9
8	CKL175755	2.18***	1.90***	0.02	-0.56***	1.00***	3.9
9	CKL175798	-0.97***	-1.77***	-0.00	0.07	0.37***	4.2
10	CKL176082	1.99***	1.96***	0.06**	-0.03	0.34***	3.6
11	CKDHL120918	-2.22***	-3.14***	-0.04*	0.14*	0.31***	3.6
12	CML585	-2.54***	-1.61***	-0.05*	0.62***	-0.79***	3.3
13	CKL14546	-1.27***	-1.27***	0.04*	0.90***	-0.27***	2.5
14	CML444	2.36***	1.80***	-0.01	0.03	-0.29***	3.3
15	***KS23-6***[†]						***3.6***
SE/LSD$_{(0.05)}$[$]		0.13	0.19	0.02	0.12	0.06	0.8

*, **, and *** indicate significance at the 0.05, 0.01, and 0.001 levels, respectively.
[†]KS23-6, an MLN resistant donor line was used as a check in the line trial.
[$]SE of GCA effects; LSD for *per se* performance.
DTA, days to anthesis; DTS, days to silking; GY, grain yield; EPP, ears per plant; TLB, Turcicum leaf blight.

development of MLN resistant parental inbred lines with favorable alleles for key agronomic and adaptive traits is critical for the identification of high-yielding adapted MLN-resistant maize varieties. Our objective was to understand the genetics of resistance to MLN and other key traits in the late maturity maize germplasm, information that can be utilized in designing a breeding strategy for MLN resistance. The lines used in this study differed in terms of their source germplasm (drought tolerance, adaptation, and MLN resistance), and selection history with some lines developed through pedigree breeding and others developed through the DH technique. In the present study, highly significant genotypic differences for MLN disease resistance parameters and agronomic traits under artificial inoculation conditions were revealed. Highly significant genotypic differences were detected for all traits except EPP under rainfed conditions. These findings indicate that there was sufficient genetic variability for most of the traits studied in this set of germplasm, implying that progress from selection and ultimately genetic gain can be made in breeding for MLN resistance and other adaptive traits. Access to diverse germplasm resources has enabled CIMMYT to make progress in the development of MLN-resistant germplasm to combat the disease in Eastern Africa (Prasanna et al., 2020).

The broad-sense heritability for MLN disease resistance parameters was high (H^2 = 0.92–0.95) in this study which suggests that a greater proportion of the observed phenotypic variation in this germplasm was due to genetic variability, and that breeding for MLN resistance can lead to considerably high genetic gains from selection for these traits. This conclusion is supported by the moderately high narrow-sense heritability (h^2 = 0.52–0.56) for MLN disease resistance parameters which suggested that more than 50% of the genetic control was attributed to additive genetic effects. The implication is that the resistance of this germplasm to MLN can be improved relatively rapidly. There are no published reports on narrow-sense heritability for MLN disease resistance parameters to compare our results with. High broad-sense heritability estimates for virus diseases of maize have been previously reported, e.g., for MLN (Beyene et al., 2017), MCMV (Jones et al., 2018), and SCMV (Pokorny and Porubova, 2006).

The genetics of virus diseases have been investigated in tropical maize using a variety of intermediate maturity maize germplasm (Mutengwa et al., 2012; Nyaligwa et al., 2017; Beyene et al., 2017; Nyaga et al., 2020). In our study, we used maize in the late-maturity category to investigate the combining abilities and reciprocal effects in this germplasm. The results of this study revealed a greater contribution of GCA variance than of SCA variance for all traits under artificial MLN inoculation and for agronomic and disease traits under rainfed conditions. This suggests a preponderance of additive genetic effects in the inheritance of these traits in this germplasm. This result is consistent with findings for MLN in a study under artificial inoculation (Beyene et al., 2017), for maize

10.3389/fpls.2024.1506139

TABLE 7 Mean squares from ANOVA of the 14-parent diallel for MLN disease resistance parameters under artificial inoculation with MLN based on Hayman's (1954a) method.

Item	Effect	df	MLN1	MLN2	MLN3	MLN4
a	Additivity	13	32.78***	50.59***	66.48***	68.84***
b	Dominance	91	3.56***	5.56***	6.42***	7.06***
b1	Directional dominance	1	3.01***	8.92***	8.28***	7.72***
b2	Gene distribution	13	2.59***	4.98***	7.18***	9.08***
b3	Residual dominance	77	3.73***	5.61***	6.26***	6.71***
c	Maternal	13	0.13ns†	0.20ns	0.18ns	0.18ns
d	Reciprocal	78	0.18ns	0.19ns	0.24ns	0.23ns
a × Environment (E)		39	0.75***	0.65***	0.49***	0.71***
b × E		273	0.38***	0.33***	0.34***	0.46***
b1 × E		3	10.77***	10.96***	10.38***	12.71***
b2 × E		39	0.56***	0.43***	0.38***	0.89***
b3 × E		231	0.22***	0.17ns	0.20ns	0.24ns
c × E		39	0.29***	0.34***	0.52***	0.60***
d × E		234	0.17***	0.13***	0.18***	0.16***
Error		774	0.17	0.17	0.22	0.23

*** indicates significance at the 0.001 level.
†ns, not significant.
MLN1, MLN2, MLN3, and MLN4 are maize lethal necrosis disease ratings at 21, 28, 35, and 42 days after inoculation, respectively.

streak virus (Mutengwa et al., 2012; Kim et al., 1989) and for maize dwarf mosaic virus (Loesch and Zuber, 1972; Rosenkranz and Scott, 1987). Both additive and dominant genes have been reported to condition virus resistance in maize (Naidu and Josephson, 1976; Zambrano et al., 2014). With a predominance of GCA over SCA, early testing may be more effective, and promising hybrids can be identified and selected mainly based on the prediction from GCA effects and the most resistant hybrids can be obtained by crossing the parents with the highest GCA (Baker, 1978; Makumbi et al., 2011). To improve both MLN resistance and grain yield potential, breeders can select lines with significant negative GCA effects for MLN resistance, significant positive GCA effects for grain yield, and

TABLE 8 Mean genetic parameters for MLN disease resistance parameters based on diallel analysis of 14 inbred lines evaluated under artificial inoculation with MLN at Naivasha over 3 years (2020–2022).

Genetic parameter	MLN1	MLN2	MLN3	MLN4
Additive variance component (D)	1.17** ± 0.11	1.94** ± 0.14	2.84** ± 0.24	2.94** ± 0.31
Dominant variance component (H_1)	0.93** ± 0.20	1.68** ± 0.26	2.06** ± 0.46	2.32** ± 0.60
Dominant variance component (H_2)	0.72** ± 0.16	1.22** ± 0.22	1.39** ± 0.38	1.52** ± 0.49
Dominant effect component (h^2)	0.08 ± 0.11	0.27 ± 0.14	0.25 ± 0.25	0.20 ± 0.32
Relative frequency of dominant and recessive alleles (F)	0.80** ± 0.24	1.52** ± 0.31	2.34** ± 0.55	2.52** ± 0.70
Environmental variation (E)	0.08 ± 0.03	0.08 ± 0.04	0.11 ± 0.06	0.11 ± 0.08
Mean degree of dominance ($\sqrt{H_1/D}$)	0.89	0.93	0.85	0.89
Minimum number of groups of genes	3.53	3.82	3.63	3.31
Correlation (*r*) between Wr+Vr and Yr	-0.59	-0.61	-0.59	-0.64
Broad-sense heritability	0.85	0.90	0.89	0.89
Narrow-sense heritability	0.52	0.53	0.56	0.55

** indicates significance at the 0.01 level.
MLN1, MLN2, MLN3, and MLN4 are maize lethal necrosis disease ratings at 21, 28, 35, and 42 days after inoculation, respectively.

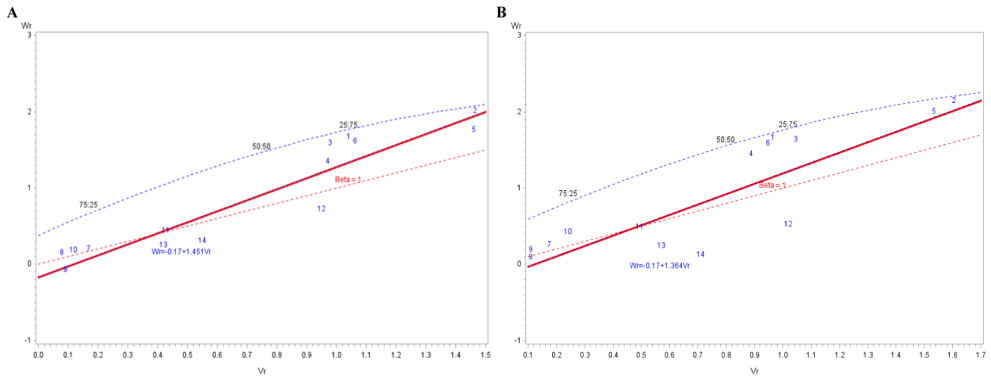


FIGURE 2
The relationship between the variance of the F_1 for each parental line (*Vr*) and their covariance with the non-recurrent parent (*Wr*) for the MLN3 **(A)** and MLN4 **(B)** disease severity scores across four seasons, 2020–2022.

significant negative GCA effects for foliar diseases to create new hybrid combinations for testing in the relevant target environments. Performance of hybrids can also be predicted based on the performance of single crosses (e.g. Zuber et al., 1973). Recurrent selection that emphasizes GCA can be an effective strategy to improve MLN resistance in this germplasm, as has been recommended for other virus diseases of maize (Josephson and Naidu, 1971; Kim et al., 1989; Mutengwa et al., 2012). With the implementation of forward breeding for MLN and other diseases such as MSV, the identification of MLN resistant lines that possess other adaptive traits for hybrid development should lead to faster genetic gains.

The consistency observed in the GCA effects across the four time points suggests rating genotypes for MLN resistance on multiple dates may not be necessary. Evaluating genotypes at 35 days (MLN3) and 42 (MLN4) days after inoculation should be sufficient for accurate assessment of resistance. The results revealed that five inbred lines had consistently significant desirable (negative for reduced disease) GCA effects for the four MLN resistance scores and AUDPC. This suggested that these lines have favorable alleles for resistance to MLN and can slow disease progression. Inbred line 6 (CKL176616) which had the largest desirable GCA effects for the four MLN resistance ratings derived its resistance alleles from KS23-5 unlike three of the four lines (CKL181281, CKL181379, and CKL182037) whose resistance was from KS23-6. Interestingly, among the five inbred lines, inbred line CKL175755 is not known to have any pedigree breeding history of resistance to viruses. Two other lines (CKL175798 and CKL176082), also without a background of virus resistance showed significant negative GCA effects for MLN4. Detailed studies (e.g., Jones et al., 2018) under artificial MLN, MCMV, and SCMV inoculation in net houses are needed to confirm the response of these lines as they may offer an additional source of alleles for MLN resistance. To date, only lines KS23-5 and KS23-6 have validated QTLs for MLN resistance (Murithi et al., 2021; Awata et al., 2021), but QTLs for resistance to MCMV, one of the two viruses that cause MLN have been mapped in other lines (Jones et al., 2018). In our study, the inbred line CKDHL120918 had significant positive GCA effects for MLN but this line was reported to have significant negative GCA effects for MLN in an earlier study (Beyene et al., 2017). The differences in GCA effects of this line between the two studies could be due to variation in the diallel method used, as different methods can impact GCA estimates (Fan et al., 2014). We used Method 3 in this study, while Beyene et al. (2017) used Method 4 of Griffing (1956). Four inbred lines (CKL181281, CKL182037, CKL176616, and CKL175798) out of the five that had significant negative GCA effects for all MLN disease resistance parameters and AUDPC also expressed significant positive GCA effects for GY under artificial MLN inoculation. These findings demonstrate progress made in developing lines with favorable alleles for MLN-resistance (reduced disease susceptibility) while also combining beneficial alleles for GY under both MLN-infected and rainfed conditions. Furthermore, two of the four lines (CKL181281 and CKL182037) had significant positive GCA effects for GY and significant negative GCA effects for *Turcicum* leaf blight under rainfed conditions. These two lines with desirable GCA effects for GY and disease resistance across different conditions have the potential as inbred line testers for the MLN breeding program if the right testing strategies (e.g. Castellanos et al., 1998; Pswarayi and Vivek, 2008) are used to confirm their suitability as testers. Suitable testers must correctly classify and discriminate efficiently among test entries (Rawlings and Thompson, 1962).

The use of reciprocal crosses provides a quantitative method to assess the contribution of maternal effects in the inheritance of a trait. When reciprocal differences are strong, parental inbred performance affects the choice of a female parent in a hybrid (Mann and Pollmer, 1981). In this study, reciprocal and maternal effects were not significant for MLN disease resistance parameters, indicating that the disease parameters recorded were not influenced by maternal effects or cytoplasmic inheritance. This suggests that a line that is resistant to MLN can be used either as a female or male parent in a hybrid combination, although other traits such as seed

producibility and pollen production must be considered. According to Roach and Wullf (1987), reciprocal crosses have similar nuclear genetic contributions, and any divergence in the performance of reciprocal pairs is due to a maternal or perhaps a paternal effect. The absence of maternal influence on the MLN disease resistance parameters signifies the predominance of additive gene action for MLN resistance since maternal effects are assumed to result from nonadditive gene action. Furthermore, maternal effects can potentially decrease the accuracy of genetic studies. Both cytoplasmic and nuclear maternal genetic effects may increase the observed genetic variance, but if the trait is fully controlled by maternal factors, they could curtail the response to selection (Roach and Wullf, 1987). Therefore, lack of maternal effects in this study suggests that the response to selection for MLN resistance will be minimally impacted. Our results indicated significant reciprocal effects for several agronomic traits including DTA and GY under MLN inoculation, and DTA, EPP, and GY under rainfed conditions. These findings were consistent with other reports of significant reciprocal effects for GY in maize (Ordás et al., 2008; Kagoda et al., 2011). However, our results contrast with those of Jumbo and Carena (2008) and Machida et al. (2010) who reported no significant effects for GY.

The development of multiple stress tolerant maize germplasm is an objective of many breeding programs in SSA, and this requires the selection of parental lines with suitable breeding values for the target traits. In SSA, stress tolerances including tolerance to foliar diseases (TLB and gray leaf spot), viral diseases (MSV and MLN) and low soil fertility and drought stress are required in certain combinations in hybrids for commercial production. The results showed that several lines had desirable GCA effects for a combination of some of the stresses. These lines should be tested in hybrid combinations under managed abiotic stress conditions mainly drought and low soil fertility following established protocols (e.g. Njeri et al., 2017; Makumbi et al., 2018b). Lines that have the highest breeding values across several stresses can be used as parents in biparental populations for development of new inbred lines. Further improvement of tropical germplasm using the new lines can be achieved with the application of modern tools and techniques such as DH, genomic selection, and marker-assisted selection with improvements in phenotyping methods to increase the rate of genetic gain and develop new multiple stress tolerant inbred lines (Cooper et al., 2014; Cobb et al., 2019; Prasanna et al., 2021; 2022).

The diallel analysis model of Hayman (1954a, b) provides estimates of several quantitative genetic parameters for the traits of interest. This method of analysis has been applied to the diallel analysis of quantitative traits alongside Griffing (1956) method to gain a deeper understanding of the inheritance of a trait beyond what a single method of diallel analysis can provide (e.g. Naidu and Josephson, 1976; Hamid et al., 1982; Betrán et al., 2003; Kagoda et al., 2011). The current study is the first to investigate the quantitative genetic parameters of MLN resistance using Hayman's model. The results indicated that additive effects had a greater contribution in the inheritance of MLN resistance compared to dominance effects based on Hayman's method, a result similar to that obtained using Griffing's method for combining ability analysis. A study on the inheritance of *Helminthosporium* leaf spot in maize (Hamid et al., 1982) also reported congruence between results from Hayman's and Griffing's models for several disease resistance parameters. Our results revealed that the alleles controlling resistance or susceptibility to MLN were not equally distributed among the lines used in this study ($H_2 < H_1$). This was evident through the distribution of the parental arrays along the regression line on the W_r-V_r graph. The 14 parents were therefore unique with respect to the dominance and/or epistatic effects of the genes they possess (Allard, 1956). The W_r-V_r graphs for MLN3 and MLN4 were similar in the placement of the arrays along the regression line. The group of lines that were among the most resistant to MLN (P2 [CKL181281] and P5 [CKL182037]) were located far from the origin at the upper end of the regression line, which indicated that the alleles conditioning resistance in these lines were mostly recessive. These inbred lines had desirable GCA effects for MLN resistance and AUDPC. A group of four lines (P7, P8, P9, and P10) was located closer to the origin of the regression line, which indicated that this set of lines possessed more dominant alleles. Some of these lines, especially P8 and P9, showed significant negative GCA effects for MLN and AUDPC. This finding is of particular interest and necessitates further investigation to understand the genetic basis of resistance in these lines given that resistance to MLN has been reported to be recessive. The positioning of P8 and P9 with respect to other lines with significant negative GCA effects for MLN and exhibited resistance such as P2, P3, P5 and P6, was surprising. This may suggest that these lines exhibited similar phenotypes through different genetic mechanisms (Luckett, 1989). These lines should be tested under artificial inoculation with individual MCMV and SCMV isolates in a net house for better discrimination in terms of resistance to these viruses.

5 Conclusions

This study revealed that additive genetic effects contribute significantly to the inheritance of MLN in late maturity germplasm adapted to eastern Africa. Five inbred lines, three of which were derived from introgression of MLN resistance from KS23-5 and KS23-6 exhibited significant desirable GCA effects for MLN resistance and GY under artificial MLN conditions. Three of these lines also showed significant desirable GCA effects for GY under disease and rainfed conditions and have the potential to contribute to the development of multiple stress tolerant hybrids for the target product profile. We identified four inbred lines with desirable GCA effects for MLN resistance despite having no known breeding history of virus resistance. Detailed studies under artificial inoculation with individual viruses (MCMV and SCMV) and MLN should help to decipher the genetic basis of resistance to MLN in this group of lines. Reciprocal effects were of minor importance; therefore, breeding programs can use any MLN resistant inbred line

as a female or male in hybrid combinations without a significant effect on the MLN response in the final product. The graphical method of analysis revealed the distribution of the lines in relation to the abundance of recessive or dominant alleles, and this information will be useful for selecting parents for biparental populations and hybrid development.

Data availability statement

The data sets used in this study have all been analyzed and summarized in the Supplementary Tables presented.

Author contributions

HK: Data curation, Formal analysis, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. DM: Investigation, Methodology, Software, Writing – original draft, Writing – review & editing, Conceptualization. FN: Investigation, Supervision, Writing – review & editing. VW: Methodology, Validation, Writing – review & editing. LS: Data curation, Investigation, Supervision, Writing – review & editing. WM: Investigation, Supervision, Writing – review & editing, Software. MG: Data curation, Formal analysis, Writing – review & editing. BP: Funding acquisition, Project administration, Resources, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was supported by the Bill and Melinda Gates Foundation (BMGF), Foundation for Food and Agriculture Research (FFAR), and the United States Agency for International Development (USAID) through the Accelerating Genetic Gains in Maize and Wheat for Improved Livelihoods project (INV–003439).

Acknowledgments

We thank Wilson Mwaura and James Mutinde at the CIMMYT-KALRO MLN Screening Facility at Naivasha, and Gabriel Ambani and Christine Litali at the KALRO Kakamega Non-Ruminant Research Center for their assistance with trial management and data collection. We also thank Joseph S. Kasango for his assistance with data curation.

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

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2024.1506139/full#supplementary-material

References

Adams, I. P., Harju, V., Hodges, T., Hany, U., Skelton, A., Rai, S., et al. (2014). First report of maize lethal necrosis disease in Rwanda. *New Dis. Rep.* 29, 22. doi: 10.5197/j.2044-0588.2014.029.022

Allard, R. W. (1956). The analysis of genetic-environmental interactions by means of diallel crosses. *Genetics* 41, 305–318. doi: 10.1093/genetics/41.3.305

Alvarado, G., Rodríguez, F. M., Pacheco, A., Burgueño, J., Crossa, J., Vargas, M., et al. (2020). META-R: A software to analyze data from multi-environment plant breeding trials. *Crop J.* 8, 745–756. doi: 10.1016/j.cj.2020.03.010

Asea, G., Bigirwa, G., Adipala, E., Powera, S. A., Pratt, R. C., and Lipps, P. E. (2002). Effect of *Cercospora zeae-maydis* infested maize residue on the progress and spread of gray leaf spot of maize in central Uganda. *Ann. Appl. Biol.* 140, 177–185. doi: 10.1111/j.1744-7348.2002.tb00171.x

Awata, L. A. O., Ifie, B. E., Danquah, E., Jumbo, M. B., Suresh, L. M., Gowda, M., et al. (2021). Introgression of maize lethal necrosis resistance quantitative trait loci into susceptible maize populations and validation of the resistance under field conditions in Naivasha, Kenya. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.649308

Baker, R. J. (1978). Issues in diallel analysis. *Crop Sci.* 18, 533–536. doi: 10.2135/cropsci1978.0011183X001800040001x

Betrán, F. J., Beck, D., Bänziger, M., and Edmeades, G. O. (2003). Genetic analysis of inbred and hybrid yield under stress and nonstress environments in tropical maize. *Crop Sci.* 43, 807–817. doi: doi:10.2135/cropsci2003.8070

Beyene, Y., Gowda, M., Mugo, S., Olsen, M., Oikeh, S. O., Juma, C., et al. (2017). Genetic analysis of tropical maize inbred lines for resistance to maize lethal necrosis disease. *Euphytica* 213, 224. doi: 10.1007/s10681-017-2012-3

Braidwood, L., Quito-Avila, D. F., Cabanas, D., Bressan, A., Wangai, A., and Baulcombe, D. C. (2018). Maize chlorotic mottle virus exhibits low divergence between differentiated regional subpopulations. *Sci. Rep.* 8, 1173. doi: 10.1038/s41598-018-19607-4

Brewbaker, J. L. (2009). Registration of nine maize populations resistant to tropical diseases. *J. Plant Registrations* 3, 10–13. doi: 10.3198/jpr2008.07.0396crc

Castellanos, J. S., Hallauer, A. R., and Cordova, H. S. (1998). Relative performance of testers to identify elite lines of corn (*Zea mays* L.). *Maydica* 43, 217–226.

Cobb, J. N., Juma, R. U., Biswas, P. S., Arbelaez, J. D., Rutkoski, J., Atlin, G., et al. (2019). Enhancing the rate of genetic gain in public–sector plant breeding programs: lessons from the breeder's equation. *Theor. Appl. Genet.* 132, 627–645. doi: 10.1007/s00122-019-03317-0

Cooper, M., Messina, C. D., Podlich, D., Radu Totir, L., Baumgarten, A., Hausmann, N. J., et al. (2014). Predicting the future of plant breeding: complementing empirical evaluation with genetic prediction. *Crop Pasture Sci.* 65, 311–336. doi: 10.1071/CP14007

De Groote, H., Munyua, B. G., Palmas, S., Suresh, L. M., Bruce, A. Y., and Kimenju, S. (2021). Using panel community surveys to track the impact of crop pests over time and space – the case of maize lethal necrosis (MLN) disease in Kenya from 2013 to 2018. *Plant Dis.* 105, 1259–1271. doi: 10.1094/PDIS-08-20-1730-SR

De Groote, H., Oloo, F., Tongruksawattana, S., and Das, B. (2016). Community-survey based assessment of the geographic distribution and impact of maize lethal necrosis (MLN) disease in Kenya. *Crop Prot.* 82, 30–35. doi: 10.1016/j.cropro.2015.12.003

Erenstein, O., Jaleta, M., Sonder, K., Mottaleb, K., and Prasanna, B. M. (2022). Global maize production, consumption and trade: trends and RandD implications. *Food Secur.* 14, 1295–1319. doi: 10.1007/s12571-022-01288-7

Fan, X. M., Zhang, Y. D., Yao, W. H., Bi, Y. Q., Liu, L., Chen, H. M., et al. (2014). Reciprocal diallel crosses impact combining ability, variance estimation, and heterotic group classification. *Crop Sci.* 54, 89–97. doi: 10.2135/cropsci2013.06.0393

FAOSTAT (2021). *Food and Agricultural Organization statistical database* (Rome, Italy: FAO). Available at: http://faostat.fao.org.

Gowda, M., Beyene, Y., Makumbi, D., Semagn, K., Olsen, M., Bright, J. M., et al. (2018). Discovery and validation of genomic regions associated with resistance to maize lethal necrosis in four biparental populations. *Mol. Breed.* 38, 66. doi: 10.1007/s11032-018-0829-7

Griffing, B. (1956). Concept of general and specific combining ability in relation to diallel crossing systems. *Aust. J. Biol. Sci.* 9, 463–493. doi: 10.1071/BI9560463

Hallauer, A. R., Carena, M., and Miranda Filho, J. B. (2010). *Quantitative genetics in maize breeding. 3rd ed* (Ames, IA: Iowa State Univ. Press).

Hamid, A. H., Ayers, J. E., and Hill, Jr. R.R. (1982). The inheritance of resistance to race 3 of *Cochliobolus carbonum*. *Phytopathology* 72, 1173–1177. doi: 10.1094/Phyto-72-1173

Hayman, B. I. (1954a). The analysis of variance of diallel tables. *Biometrics* 10, 235–244. doi: 10.2307/3001877

Hayman, B. I. (1954b). The theory and analysis of diallel crosses. *Genetics* 39, 789–809. doi: 10.1093/genetics/39.6.789

Jampatong, S., Thung-Ngean, M., Balla, C., Boonrumpun, P., Mekarun, A., Jompuk, C., et al. (2010). Evaluation of improved maize populations and their diallel crosses for yield. *Kasetsart Jounal (Natural Science)* 44, 523–528.

Jones, M. W., Penning, B. W., Jamann, T. M., Glaubitz, J. C., Romay, C., Buckler, E. S., et al. (2018). Diverse chromosomal locations of quantitative trait loci for tolerance to maize chlorotic mottle virus in five maize populations. *Phytopathology* 108, 748–758. doi: 10.1094/PHYTO-09-17-0321-R

Josephson, L. M., and Naidu, B. (1971). Reaction in diallel crosses of corn inbreds (*Zea mays* L.) to maize dwarf mosaic virus. *Crop Sci.* 11, 664–667. doi: 10.2135/cropsci1971.0011183X001100050018x

Jumbo, M. B., and Carena, M. J. (2008). Combining ability, maternal, and reciprocal effects of elite early-maturing maize population hybrids. *Euphytica* 162, 325–333. doi: doi 10.1007/s10681-007-9618-9

Kagoda, F., Derera, J., Tongoona, P., Coyne, D. L., and Lorenzen, J. (2011). Genetic analysis of resistance to nematodes in inbred maize (*Zea mays* L.) and maize hybrids. *Euphytica* 182, 377–393. doi: 10.1007/s10681-011-0512-0

Kim, S. K., Efron, Y., Fajemisin, J. M., and Buddenhagen, I. W. (1989). Mode of gene action for resistance in maize to maize streak virus. *Crop Sci.* 29, 890–894. doi: 10.2135/cropsci1989.0011183X002900040009x

Kulkarni, H. Y. (1973). Notes on East African plant virus diseases: 5. Identification and economic importance of sugar-cane mosaic virus in maize in East Africa. *East Afr. Agric. Forestry J.* 39, 158–164. doi: 10.1080/00128325.1973.11662633

Kyetere, D. T., Ming, R., McMullen, M. D., Pratt, R. C., Brewbaker, J., and Musket, T. (1999). Genetic analysis of tolerance to maize streak virus in maize. *Genome* 42, 20–26. doi: 10.1139/g98-099

Loesch, J. P.J., and Zuber, M. S. (1972). Inheritance of resistance to maize dwarf mosaic virus. *Crop Sci.* 12, 350–352. doi: 10.2135/cropsci1972.0011183X001200030027x

Louie, R. (1980). Sugarcane mosaic virus in Kenya. *Plant Dis.* 64, 944–947. doi: 10.1094/PD-64-944

Luckett, D. J. (1989). Diallel analysis of yield components, fibre quality and bacterial blight resistance using spaced plants of cotton. *Euphytica* 44, 11–21. doi: 10.1007/BF00022594

Lukanda, M., Owati, A., Ogunsanya, P., Valimunzigha, K., Katsongo, K., Ndemere, H., et al. (2014). First report of maize chlorotic mottle virus infecting maize in the Democratic Republic of Congo. *Plant Dis.* 98, 1448. doi: 10.1094/PDIS-05-14-0484-PDN

Machida, L., Derera, J., Tongoona, P., and MacRobert, J. (2010). Combining ability and reciprocal cross effects of elite quality protein maize inbred lines in subtropical environments. *Crop Sci.* 50, 1708–1717. doi: 10.2135/cropsci2009.09.0538

Mahuku, G., Lockhart, B. E., Wanjala, B., Jones, M. W., Kimunye, J. N., Stewart, L. R., et al. (2015a). Maize lethal necrosis (MLN), an emerging threat to maize-based food security in sub-Saharan Africa. *Phytopathology* 105, 956–965. doi: 10.1094/PHYTO-12-14-0367-FI

Mahuku, G., Wangai, A., Sadessa, K., Teklewold, A., Wegary, D., Ayalneh, D., et al. (2015b). First report of maize chlorotic mottle virus and maize lethal necrosis on maize in Ethiopia. *Plant Dis.* 99, 1870. doi: 10.1094/PDIS-04-15-0373-PDN

Makumbi, D., Alvarado, G., Crossa, J., and Burgueño, J. (2018a). SASHAYDIALL: A SAS program for Hayman's diallel analysis. *Crop Sci.* 58, 1605–1615. doi: 10.2135/cropsci2018.01.0047

Makumbi, D., Assanga, S., Diallo, A., Magorokosho, C., Asea, G., Worku, M., et al. (2018b). Genetic analysis of tropical midaltitude-adapted maize populations under stress and nonstress conditions. *Crop Sci.* 58, 1492–1507. doi: 10.2135/cropsci2017.09.0531

Makumbi, D., Betrán, F. J., Bänziger, M., and Ribaut, J. M. (2011). Combining ability, heterosis and genetic diversity in tropical maize (*Zea mays* L.) under stress and non-stress conditions. *Euphytica* 180, 143–162. doi: 10.1007/s10681-010-0334-5

Mann, C. E., and Pollmer, W. G. (1981). Reciprocal-cross differences between maize hybrids of inbred lines from different gene pools. *Maydica* 26, 263–271.

Marenya, P. P., Erenstein, O., Prasanna, B., Makumbi, D., Jumbo, M., and Beyene, Y. (2018). Maize lethal necrosis disease: Evaluating agronomic and genetic control strategies for Ethiopia and Kenya. *Agric. Syst.* 162, 220–228. doi: 10.1016/j.agsy.2018.01.016

Mather, K., and Jinks, J. L. (1971). *Biometrical genetics. 2nd Edition* (London: Chapman and Hall).

Menkir, A., and Ayodele, M. (2005). Genetic analysis of resistance to grey leaf spot of midaltitude maize inbred lines. *Crop Sci.* 45, 163–170. doi: 10.2135/cropsci2005.0803

Menkir, A., Makumbi, D., and Franco, J. (2012). Assessment of reaction patterns of hybrids to Striga hermonthica (Del.) Benth. under artificial infestation in Kenya and Nigeria. *Crop Sci.* 52, 2528–2537. doi: 10.2135/cropsci2012.05.0307

Mudde, B., Olubayo, F., Miano, D., Asea, G., Kilalo, D., Kiggundu, A., et al. (2018). Distribution, incidence, and severity of maize lethal necrosis disease in major maize growing agro-ecological zones of Uganda. *J. Agric. Sci.* 10, 72–85. doi: 10.5539/jas.v10n6p72

Murithi, A., Olsen, M. S., Kwemoi, D. B., Ogugo, V., Ertiro, B. T., Suresh, L. M., et al. (2021). Discovery and validation of a recessively inherited major-effect QTL conferring resistance to maize lethal necrosis (MLN) disease. *Front. Genet.* 12767883. doi: 10.3389/fgene.2021.767883

Mutengwa, C. S., Gandiwa, N., and Muchena, S. C. (2012). Genetic analysis of resistance to maize streak virus disease in dwarf maize germplasm. *Afr. J. Agric. Res.* 7, 6456–6460. doi: 10.5897/AJAR11.2357

Mwatuni, F. M., Nyende, A. B., Njuguna, J., Zhonguo, X., Machuka, E., and Stomeo, F. (2020). Occurrence, genetic diversity, and recombination of maize lethal necrosis disease-causing viruses in Kenya. *Virus Res.* 286, 198081. doi: 10.1016/j.virusres.2020.198081

Naidu, B., and Josephson, L. M. (1976). Genetic analysis of resistance to the corn virus disease complex. *Crop Sci.* 16, 167–172. doi: 10.2135/cropsci1976.0011183X001600020001x

Niblett, C. L., and Claflin, L. E. (1978). Corn lethal necrosis - a new virus disease of corn in Kansas. *Plant Dis. Rep.* 62, 15–19.

Njeri, S. G., Makumbi, D., Warburton, M. L., Diallo, A., Jumbo, M. B., and Chemining'wa, G. (2017). Genetic analysis of tropical quality protein maize (*Zea mays* L.) germplasm. *Euphytica* 213, 261. doi: 10.1007/s10681-017-2048-4

Nyaga, C., Gowda, M., Beyene, Y., Murithi, T. W., Burgueño, J., Toledo, F., et al. (2020). Hybrid breeding for MLN resistance: Heterosis, combining ability, and hybrid prediction. *Plants* 9, 468. doi: 10.3390/plants9040468

Nyaligwa, M. L., Shimelis, H., Laing, D. M., and Mwadzingeni, L. (2017). Combining ability for grain yield and resistance to maize streak virus in maize. *Maydica* 62, 1–7.

Ordás, B., Malvar, R. A., Ordás, A., and Revilla, P. (2008). Reciprocal differences in sugary × sugary enhancer sweet corn hybrids. *J. Am. Soc. Hortic. Sci.* 133, 777–782. doi: 10.21273/JASHS.133.6.777

Patterson, H. D., and Williams, E. R. (1976). A new class of resolvable incomplete block designs. *Biometrika* 63, 83–89. doi: 10.1093/biomet/63.1.83

Pokorny, R., and Porubova, M. (2006). Heritability of resistance in maize to the Czech isolate of sugarcane mosaic virus. *Cereal Res. Commun.* 34, 1081–1086. doi: 10.1556/CRC.34.2006.2-3.241

Prasanna, B. M. (Ed.) (2021). *Maize Lethal Necrosis (MLN): A Technical manual for disease management* (CDMX: CIMMYT, Mexico). Available at: https://hdl.handle.net/10883/21703.

Prasanna, B. M., Burgueño, J., Beyene, Y., Makumbi, D., Asea, G., Woyengo, V., et al. (2022). Genetic trends in CIMMYT's tropical maize breeding pipelines. *Sci. Rep.* 12, 20110. doi: 10.1038/s41598-022-24536-4

Prasanna, B. M., Cairns, J. E., Zaidi, P. H., Beyene, Y., Makumbi, D., Gowda, M., et al. (2021). Beat the stress: breeding for climate resilience in maize for the tropical rainfed environments. *Theor. Appl. Genet.* 134, 1729–1752. doi: 10.1007/s00122-021-03773-7

Prasanna, B. M., Suresh, L. M., Mwatuni, F., Beyene, Y., Makumbi, D., Gowda, M., et al. (2020). Maize lethal necrosis (MLN): Efforts toward containing the spread and impact of a devastating transboundary disease in sub-Saharan Africa. *Virus Res.* 282, 197943. doi: 10.1016/j.viruses.2020.197943

Pswarayi, A., and Vivek, B. S. (2008). Combining ability amongst CIMMYT's early maturing maize (*Zea mays* L.) germplasm under stress and non-stress conditions and identification of testers. *Euphytica* 162, 353–362. doi: 10.1007/s10681-007-9525-0

Rawlings, J. O., and Thompson, D. L. (1962). Performance level as criterion for the choice of maize testers. *Crop Sci.* 2, 217–220. doi: 10.2135/cropsci1962.0011183X000200030012x

Redinbaugh, M. G., and Stewart, L. R. (2018). Maize lethal necrosis: an emerging, synergistic viral disease. *Annu. Rev. Virol.* 5, 301–322. doi: 10.1146/annurev-virology-092917-043413

Roach, D. A., and Wullf, R. (1987). Maternal effects in plants: evidence and ecological and evolutionary significance. *Annu. Rev. Ecology Evol. Systematics* 18, 209–235. doi: 10.1146/annurev.es.18.110187.001233

Rodríguez, F., Alvarado, G., Pacheco, A., Crossa, J., and Burgueno, J. (2020). AGD-R (Analysis of Genetic Designs with R for Windows). Version 5.0. CIMMYT Research and Data Software Repository. Available online at: https://hdl.handle.net/11529/10202 (Accessed 11th October 2023).

Rosenkranz, E., and Scott, G. E. (1987). Type of gene action in the resistance to maize chlorotic dwarf virus in corn. *Phytopathology* 77, 1293–1296. doi: 10.1094/phyto-77-1293

SAS Institute (2016). *SAS/STAT 9.4 User's guide* (Cary, NC: SAS Institute Inc.).

Shaner, G., and Finney, R. E. (1977). The effect of nitrogen fertilization on the expression of slow-mildewing resistance in Knox wheat. *Phytopathology* 67, 1051–1056. doi: 10.1094/Phyto-67-1051

Shiferaw, B., Prasanna, B. M., Hellin, J., and Bänziger, M. (2011). Crops that feed the world 6. Past successes and future challenges to the role played by maize in global food security. *Food Secur.* 3, 307–327. doi: 10.1007/s12571-011-0140-5

Sitonik, C., Suresh, L. M., Beyene, Y., Olsen, M. S., Makumbi, D., Kiplagat, O., et al. (2019). Genetic architecture of maize chlorotic mottle virus and maize lethal necrosis through GWAS, linkage analysis and genomic prediction in tropical maize germplasm. *Theor. Appl. Genet.* 132, 2381–2399. doi: 10.1007/s00122-019-03360-x

Sserumaga, J. P., Makumbi, D., Assanga, S. O., Mageto, E. K., Njeri, S. G., Jumbo, M. B., et al. (2020). Identification and diversity of tropical maize inbred lines with resistance to common rust (*Puccinia sorghi* Schwein). *Crop Sci.* 60, 2971–2989. doi: 10.1002/csc2.20345

Stewart, L. R., Willie, K., Wijeratne, S., Redinbaugh, M. G., Massawe, D., Niblett, C. L., et al. (2017). Johnsongrass mosaic virus contributes to maize lethal necrosis in East Africa. *Plant Dis.* 101, 1455–1462. doi: 10.1094/PDIS-01-17-0136-RE

Uyemoto, J. K., Bockeman, D. L., and Claflin, L. E. (1980). Severe outbreak of corn lethal necrosis disease in Kansas. *Plant Dis.* 64, 99–100. doi: 10.1094/PD-64-99

van Rensburg, G. D. J., Giliomee, J. H., and Pringle, K. L. (1991). Resistance of South African maize hybrids to maize streak virus. *South Afr. J. Plant Soil* 8, 38–42. doi: 10.1080/02571862.1991.10634577

Vivek, B. S., Odongo, O., Njuguna, J., Imanywoha, J., Bigirwa, G., Diallo, A., et al. (2010). Diallel analysis of grain yield and resistance to seven diseases of 12 African maize (*Zea mays* L.) inbred lines. *Euphytica* 172, 329–340. doi: 10.1007/s10681-009-9993-5

Wamaitha, J. M., Nigam, D., Maina, S., Stomeo, F., Wangai, A., Njuguna, J. N., et al. (2018). Metagenomic analysis of viruses associated with maize lethal necrosis in Kenya. *Virol. J.* 15, 90. doi: 10.1186/s12985-018-0999-2

Wangai, A. W., Redinbaugh, M. G., Kinyua, Z. M., Miano, D. W., Leley, P. K., Kasina, M., et al. (2012). First report of maize chlorotic mottle virus and maize lethal necrosis in Kenya. *Plant Dis.* 96, 1582. doi: 10.1094/pdis-06-12-0576-pdn

Zambrano, J. L., Jones, M. W., Brenner, E., Francis, D. M., Tomas, A., and Redinbaugh, M. G. (2014). Genetic analysis of resistance to six virus diseases in a multiple virus-resistant maize inbred line. *Theor. Appl. Genet.* 127, 867–880. doi: 10.1007/s00122-014-2263-5

Zuber, M. S., Hilderbrand, E. S., Loesch, P. J., and Keaster, A. J. (1973). Prediction of reactions to maize dwarf mosaic virus in double-cross hybrids based upon single-cross reaction. *Crop Sci.* 13, 172–175. doi: 10.2135/cropsci1973.0011183X001300020007x

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 07 March 2025
DOI 10.3389/fpls.2025.1521579

Check for updates

OPEN ACCESS

EDITED BY
Somashekhar M. Punnuri,
Fort Valley State University, United States

REVIEWED BY
Venkatesh Bollina,
Agriculture and Agri-Food Canada (AAFC), Canada
Deepanshu Jayaswal,
Indian Institute of Seed Science, India
Durga Chinthalapudi,
Mississippi State University, United States

*CORRESPONDENCE
Venkateswara R. Sripathi
v.sripathi@aamu.edu

RECEIVED 02 November 2024
ACCEPTED 31 January 2025
PUBLISHED 07 March 2025

CITATION
Karapareddy S, Anche VC, Tamatamu SR, Janga MR, Lawrence K, Nyochembeng LM, Todd A, Walker LT and Sripathi VR (2025) Profiling of rhizosphere-associated microbial communities in North Alabama soils infested with varied levels of reniform nematodes. *Front. Plant Sci.* 16:1521579. doi: 10.3389/fpls.2025.1521579

COPYRIGHT
© 2025 Karapareddy, Anche, Tamatamu, Janga, Lawrence, Nyochembeng, Todd, Walker and Sripathi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Profiling of rhizosphere-associated microbial communities in North Alabama soils infested with varied levels of reniform nematodes

Sowndarya Karapareddy[1], Varsha C. Anche[1], Sowjanya R. Tamatamu[1], Madhusudhana R. Janga[2], Kathy Lawrence[3], Leopold M. Nyochembeng[1], Antonette Todd[4], Lloyd T. Walker[1] and Venkateswara R. Sripathi[1]*

[1]College of Agricultural, Life & Natural Sciences, Alabama A&M University, Normal, AL, United States, [2]Institute of Genomics for Crop Abiotic Stress Tolerance, Department of Plant and Soil Science, Texas Tech University, Lubbock, TX, United States, [3]Department of Entomology and Plant Pathology, Auburn University, Auburn, AL, United States, [4]Department of Agriculture & Natural Resources, Delaware State University, Dover, DE, United States

Introduction: Plant roots, nematodes, and soil microorganisms have a complex interaction in the rhizosphere by exchanging or communicating through biomolecules or chemicals or signals. Some rhizospheric (including endophytic) microbes process such compounds via biogeochemical cycles to improve soil fertility, promote plant growth and development, and impart stress tolerance in plants. Some rhizospheric microbes can affect negatively on plant parasitic nematodes (PPNs) thus hindering the ability of nematodes in parasitizing the plant roots. Next-generation sequencing is one of the most widely used and cost-effective ways of determining the composition and diversity of microbiomes in such complex environmental samples.

Methods: This study employed amplicon sequencing (Illumina/NextSeq) of 16S ribosomal RNA (16S rRNA) for bacteria and Internal Transcribed Spacer (ITS2) region for fungi to profile the soil microbiome in the rhizosphere of cotton grown in North Alabama. We isolated DNA (ZymoBIOMICS) from soil samples in triplicates from four representative locations of North Alabama. Based on the level of Reniform Nematode (RN) Infestation, these locations were classified as Group A-RN Not-Detected (ND), Group B-RN Low Infestation (LI), Group C-RN Medium Infestation (MI), and Group D-RN High Infestation (HI) and determined using sieving method and microscopic examination.

Results and discussion: Our analyses identified 47,893 bacterial and 3,409 fungal Amplicon Sequence Variants (ASVs) across all groups. Among the bacterial ASVs, 12,758, 10,709, 12,153, and 11,360 unique ASVs were determined in Groups A, B, C, and D, respectively. While 663, 887, 480, and 326 unique fungal ASVs were identified in Groups A, B, C, and D, respectively. Also, the five most abundant rhizospheric bacterial genera identified were *Gaiella*, *Conexibacter*, *Bacillus*, *Blastococcus*, *Streptomyces*. Moreover, five abundant fungal genera belonging

to *Fusarium, Aspergillus, Gibberella, Cladosporium, Lactera* were identified. The tight clustering of bacterial nodes in *Actinobacteria*, *Acidobacteria*, and *Proteobacteria* shows they are highly similar and often found together. On the other hand, the close association of *Ascomycota* and *Basidiomycota* suggesting that they have different ecological roles but occupy similar niches and contribute similar functions within the microbial community. The abundant microbial communities identified in this study had a role in nutrient recycling, soil health, plant resistance to some environmental stress and pests including nematodes, and biogeochemical cycles. Our findings will aid in broadening our understanding of how microbial communities interact with crops and nematodes in the rhizosphere, influencing plant growth and pest management.

KEYWORDS

soil, rhizosphere, reniform nematode, infestation, Phyloseq, microbial diversity, bacterial and fungal communities

1 Introduction

The rhizosphere, a critical zone of soil surrounding plant roots, serves as a dynamic interface for interactions between plants and a diverse array of microorganisms. These microorganisms, including bacteria, fungi, and archaea, play vital roles in enhancing plant growth, improving soil fertility, and promoting ecosystem stability. They are involved in various processes such as nutrient cycling, organic matter decomposition, and the regulation of plant stress responses (Bais et al., 2006; Zhalnina et al., 2022). The presence of beneficial microbes in the rhizosphere can improve plant health by enhancing nutrient uptake, providing protection against pathogens, and promoting plant growth through mechanisms like nitrogen fixation and phosphorus solubilization (Lugtenberg and Kamilova, 2009; Van der Heijden et al., 2015).

In addition to their direct benefits to plant health, rhizosphere microbes also interact with plant-parasitic nematodes (PPN's), which are significant pests in agriculture. Nematodes, particularly those that feed on plant roots, cause substantial damage to crops by disrupting root function thereby affecting the plant growth that ultimately results in the yield loss. However, the rhizosphere is the niche to a wide variety of microorganisms that can influence nematode populations through several mechanisms. Beneficial bacteria and fungi in the rhizosphere can suppress nematode infestations by producing nematicidal compounds, competing for resources, or acting as biological control agents (Zhao et al., 2017; Singh et al., 2020).

Reniform nematodes (RN) is a devastating pest in agriculture due to their widespread distribution affecting several crop species and the ability to thrive in diverse soil conditions. Their infestations can significantly alter the microbial community structure within the rhizosphere, potentially leading to decreased microbial diversity and disrupted nutrient dynamics (van der Putten and Bakker, 2018). Studies have demonstrated that specific microbial taxa can improve plant health by suppressing nematode populations and enhancing nutrient availability (Latz et al., 2021). For instance, beneficial bacteria and fungi can establish symbiotic relationships with cotton roots, leading to improved nutrient uptake and overall plant vigor (Garbeva et al., 2004; Lugtenberg and Kamilova, 2009). Furthermore, these beneficial microbes in the rhizosphere can also produce bioactive compounds that directly inhibit hatching and development of nematodes (Prasad and De Vries, 2019). Identifying these microorganisms within cotton rhizospheres is crucial for developing innovative management strategies aimed at nematode control and soil nutrient enhancement, aiding in reducing the reliance on chemical pesticides (Hassan and Abo-Elyousr, 2019).

However, nematode infestations can significantly alter the structure and diversity of microbial communities in the rhizosphere (Liu et al., 2017; Naylor et al., 2021). Changes in microbial diversity, especially a reduction in beneficial bacteria and fungi, have been linked to increased nematode damage in crops such as cotton and soybean (Yuan et al., 2020a). Additionally, plant-parasitic nematodes (PPNs) can influence plant performance by altering root exudation patterns, which in turn modify the microbial composition of the rhizosphere and improve the availability of nitrogen (N) and phosphorus (P) to plants (Topalovic et al., 2020). Verschoor (2002) found that nematode feeding contributes to nutrient cycling through the excretion of ammonia (NH3), N defecation, and increased root exudation. Similarly, Xie et al. (2023) demonstrated that nematode infestations in rice altered microbial populations enhancing nutrient cycling, particularly by increasing nitrogen-fixing bacteria that support plant growth. Wang et al. (2022) reported that nematode feeding on wheat roots shifted microbial communities, favoring fungi that contribute to organic matter decomposition, thus enhancing soil nutrient availability. In another study, Patel et al. (2024) showed that nematode-induced changes in microbial diversity helped plants by promoting the activity of specific microbes involved in phosphorus cycling, supporting plant growth under nutrient-limited conditions. Increased nematode presence often correlates with a decline in

beneficial microbes, disrupting the ecological balance and negatively impacting soil health (Zhao et al., 2018; Bhattacharyya and Jha, 2012). Therefore, understanding the interplay between nematodes and microbial communities is essential for fostering sustainable agricultural practices.

The interactions between nematodes and soil microorganisms are multifaceted, encompassing competition, predation, and mutualism (Cai et al., 2023). Beneficial microbes can suppress nematode populations through antagonistic mechanisms, while nematodes may alter microbial community dynamics by changing resource availability (Gomez et al., 2019). Recent studies have emphasized the role of certain bacterial phyla, such as *Proteobacteria*, *Firmicutes*, and *Actinobacteria*, in suppressing nematode populations and promoting plant health. For example, *Proteobacteria* has been shown to produce metabolites that can inhibit nematode development, while *Firmicutes* and *Actinobacteria* contribute to enhanced plant nutrient uptake and nematode resistance (Li et al., 2022; Zhang et al., 2020). A study proposed by Naylor and Gurevitch (2021) that nematode feeding can change the composition of these microbial communities, often favoring *Ascomycota* and *Basidiomycota*, which can either help control nematode populations or shift microbial balance in ways that may reduce plant vitality. Additionally, nematodes themselves can modulate the structure of these microbial communities, causing a decline in beneficial microbes, such as those from the *Proteobacteria*, which can have cascading effects on soil health and plant resilience (Shang and Wang, 2022). Some studies have reported shifts in microbial diversity and composition in response to nematode presence, with certain taxa thriving while others diminish (De Vries and Shade, 2013). Investigating these dynamics across varying infestation levels can provide insights into how nematodes impact microbial communities and their function.

Furthermore, understanding microbial shifts in response to nematode infestations can lead to the development of targeted microbial inoculants or soil amendments that enhance beneficial microbial populations (Luo et al., 2022). These strategies offer sustainable alternatives to chemical controls, promoting long-term soil health and resilience in cotton species (Wu et al., 2019). By fostering beneficial microbial communities, it may be possible to mitigate the adverse effects of nematodes on cotton production and improve overall soil health. Advancements in molecular techniques, specifically 16S rRNA and ITS2 sequencing, have revolutionized the study of rhizosphere microbial communities. The 16S rRNA gene serves as a universal marker for bacterial identification, while the ITS2 region is widely used for characterizing fungal diversity (Ranjan et al., 2020). Together, these sequencing techniques provide a comprehensive view of the microbiome, revealing complex interactions that can influence plant health and stress responses. Incorporating R and the Phyloseq package into data analysis allows for robust profiling of microbial communities derived from sequencing studies. Phyloseq offers an efficient framework for handling and visualizing complex ecological data, enabling in-depth analysis of microbial diversity, community composition, and potential functional roles within the rhizosphere (McGuire and Triplett, 2009). This approach is particularly useful for examining the influence of RN's on microbial dynamics in cotton soils, facilitating a deeper understanding of how these interactions impact plant health and productivity.

This study aims to profile the rhizosphere microbiome of cotton soils infested with RN's across various infestation levels in North Alabama. By employing 16S rRNA and ITS2 sequencing, combined with analyses in Phyloseq, we seek to explore the intricate relationships between nematodes and microbial communities. This investigation will help identify key microbial taxa associated with different infestation levels of RN, offering insights into potential indicators of soil health and crop resilience.

2 Materials and methods

2.1 Field site selection and sample collection

The experimental design of this study primarily aimed at profiling rhizospheric microbial communities of morphometrically classified Reniform Nematode infestation levels (Nyaku et al., 2013a, b) in selected locations of North Alabama. Alabama climate is humid and subtropical geographically spread between the Gulf of Mexico at the Southern end and Appalachian Mountains at North-eastern proximity. The climatic conditions in North Alabama are uniform across these soil sample collected locations without considering the micro-climatic factors. As climatic factors and agricultural practices are relatively uniform, slight differences in soil types and the effects of soil properties on microbiome were not emphasized in our study. The soils in Jackson, Lauderdale, Madison, and Limestone counties are primarily derived from limestone and sandstone. In the selected locations, cotton is grown as monocrop or dual crop with soybean. These soils include Decatur, Dewey, Bodine, Fullerton, Madison,

TABLE 1 Geographic locations and coordinates of four counties of North Alabama.

RN Infestation Level/Groups	County Name	Location of the sample collected	Latitude	Longitude	Altitude (m)
ND-A	Jackson	Scottsboro	34.621466	-86.170195	196
LI-B	Lauderdale	Florence	34.789377	-87.746495	148
MI-C	Madison	Huntsville	34.784381	-86.505875	204
HI-D	Limestone	Belle Mina	34.661774	-86.879342	179

RN, Reniform Nematode; ND, Not-Detected; LI, Low Infestation; MI, Medium Infestation; HI, High Infestation.

Pacolet, and Cecil series, featuring textures like clayey with silt loam and sandy loam surfaces (Alabama Cooperative Extension System, 2020). Sampling locations and GPS-determined coordinates of four selected sites were outlined in Supplementary Figure 1, Table 1, respectively.

Soil samples were collected from four counties of North Alabama, USA, based on RN infestation levels: Group A - RN Not-Detected (ND), Group B - RN Low Infestation (LI), Group C - RN Medium Infestation (MI), and Group D - RN High Infestation (HI) across Jackson (ND), Lauderdale (LI), Madison (MI), and Limestone (HI), respectively. Varied levels of RN infestation and their distribution in North Alabama were determined based on our previous studies such as morphometric and DNA-based (18S and ITS) marker analyses (Nyaku et al., 2023, 2016, 2013a, 2013b), and also as reported in similar agricultural studies (Roe and Owens, 2017; Thomas et al., 2019). The infestation levels of RN in the soil samples were determined using the sieving method, where soil samples were passed through a series of sieves to isolate nematodes. First, 25 ml of soil solution with nematodes was collected from 100g of soil using a sieve method. Then, 1 ml of soil solution was aliquoted and used to count the number of nematodes under the microscope to assess morphometrically and categorize them across various infestation levels. Where, ND = 0 RN detected, LI = <2,000 RN detected, MI = 2,000-5,000 RN detected, and HI = > 5,000 RN detected. This method ensures reliable classification of the RN infestation levels, which were based on previous studies and established protocols for nematode extraction and quantification (Eisenback and Triantaphyllou, 1991; Siddiqi, 2000).

The soil sampling and collection procedures used were meticulously adhered to the Alabama Cooperative Extension System protocol (Celleti & Potter, 2006) to ensure the highest data quality for our study. Recent guidelines on soil sample handling and preservation (de la Fuente et al., 2021) were followed to minimize contamination risks and maintain microbial integrity. Rhizospheric soils were collected at a depth of approximately 10-20 cm and <12 cm from the crop using soil auger as recommended (Smith & Lee, 2023). Plant debris (including roots), stones, and other impurities were removed during the collection process. Triplicate samples of 500g for each location were collected and placed in sterile zip-lock bags. Then these samples were transported in a dark cooler with ice and stored at 4^0C until further processing (Li et al., 2023). Twelve samples collected (four counties and three replicates) were processed for nematode isolation and quantification and genomic DNA isolation. Same sample source has been used to quantify and characterize reniform nematodes for determining their levels of infestation and to isolate the DNA with higher integrity.

Nematodes were collected from the soils of Jackson, Lauderdale, Madison, and Limestone counties, morphometric measurements were made on male and female nematodes using an Olympus microscope (Olympus Optical Co. Ltd, Japan). The morphometric variables used for accurately determining the RN and their distribution in Alabama were body length, stylet length, position of vulva, spicule length, length of hyaline portion of tail, position of dorsal oesophageal gland orifice, position of excretory pore, maximum width, esophageal length and anal width. Prior to DNA extraction, the soil samples were thoroughly mixed to ensure uniformity and consistency. This step is crucial for ensuring reliable and consistent results in downstream microbiome analysis (Garcia-Sanchez et al., 2020). About 500mg of soil was measured in triplicates in 2-ml sterile microcentrifuge tubes for DNA isolation (4 x 3 = 12 samples).

2.2 Soil DNA extraction, library preparation, and sequencing

DNA was extracted from 12 soil samples using the ZymoBIOMICS-96 MagBead DNA Kit (Zymo Research, Irvine, CA), according to the manufacturer's instructions. The elution volume of DNA is 50 ul. The quantity and quality of the isolated DNA were assessed (Sambrook and Russell, 2001) using Nanodrop 1000 Spectrophotometer (Thermo Fisher Scientific, 2023), Qubit 1X dsDNA Broad Range Assay Kit (Invitrogen, 2016), and Agarose Gel Electrophoresis (Wilson, 2020), respectively. Bacterial 16S rRNA gene sequencing was conducted using the Quick-16S NGS Library Prep Kit (Zymo Research, Irvine, CA), specifically targeting the V3–V4 region of the 16S rRNA gene. Amplification was performed with designated bacterial 16S primers, adhering to the following PCR protocol: an initial denaturation step at 95°C for 3 minutes, followed by 25 cycles of 95°C for 30 seconds, 55°C for 30 seconds, and 72°C for 30 seconds, concluded with a final extension at 72°C for 5 minutes. Each sample underwent triplicate processing to enhance reproducibility (Mardis, 2008). For fungal analysis, ITS2 gene sequencing was similarly executed using the Quick-16S NGS Library Prep Kit, replacing the 16S primers with custom ITS2 primers from the Microbiome Sequencing ITS2 Primer Set. The PCR conditions for the ITS2 amplification included an initial denaturation at 95°C for 3 minutes, followed by 30 cycles of 95°C for 30 seconds, 55°C for 30 seconds, and 72°C for 30 seconds, and a final extension at 72°C for 5 minutes (White et al., 1990).

To minimize PCR chimera formation, real-time PCR monitoring was employed during library preparation for each sample. The resulting PCR products were quantified using qPCR fluorescence readings and pooled based on equal molarity. The pooled library underwent purification using the Select-a-Size DNA Clean and Concentrator (Zymo Research, Irvine, CA) and was quantified using TapeStation (Agilent Technologies, Santa Clara, CA) and Qubit 1X dsDNA High-Sensitivity Assay Kits (Thermo Fisher Scientific, Waltham, WA) (Parker et al., 2016). ZymoBIOMICS Microbial Community DNA Standards (Zymo Research, Irvine, CA) served as positive controls for each DNA extraction and targeted library preparation. Additionally, negative controls, including blank extraction and library preparation controls, were incorporated to assess the quality and potential contamination during these processes (Kozich et al., 2013). In

total, 12 libraries were sequenced on the Illumina NextSeq 2000 using a p1 (cat 20075294) reagent kit (600 cycles), with a 30% PhiX spike-in control included for sequencing (Illumina, 2019).

2.3 Bioinformatics and statistical analysis

Bioinformatics analyses were conducted to process and analyze the sequence data, starting with the improvement of read quality (Bolger et al., 2014; Chen et al., 2022). Then, the reads were paired together and assembled into genetic sequences, which were subsequently compared to reference genomes for organism identification (Li et al., 2021; Bushnell et al., 2017). The raw reads from amplicon sequencing data (16S rRNA and ITS2) were processed using the Divisive Amplicon Denoising Algorithm 2 (DADA2) pipeline in R (v4.3.2), following the procedure outlined by Callahan et al. (2016a, b). Data were then statistically analyzed with Phyloseq (v1.46.0) to create a data matrix and examine microbiome differences across and within samples. The DADA2 workflow involves quality filtering and trimming, de-replication, sequence table construction, chimera removal, taxonomy assignment, and phylogenetic tree construction. In the first step, forward reads were truncated at position 300 and reverse reads at position 200 for the 16S rRNA dataset, while for the ITS2 dataset, forward reads were truncated at position 180 and reverse reads at position 250. After being filtered by DADA2, the reads were grouped into distinct Amplicon Sequence Variants (ASVs) and aligned using the DECIPHER R package (Wright, 2015). Then, dereplication was performed to eliminate redundancy and infer ASVs without applying any arbitrary threshold, allowing for the detection of variants that differ by as little as a single nucleotide. Next, chimeras were subsequently removed using the "removeBimeraDenovo" command. Subsequently, taxonomy was assigned using the naive Bayesian classifier, employing the Ribosomal Database Project (RDP) v19 training set for 16S rRNA data (Wang et al., 2007; Cole et al., 2014) and the UNITE database v9.0 (Abarenkov et al., 2023) for ITS2 data and the phylogenetic tree was constructed with the Phangorn R package (Schliep, 2011). Finally, a Phyloseq object was used to import all the data to carry out alpha diversity, beta diversity, relative abundance with composition barplots, differential abundance analysis, heatmap, and network analyses.

Subsequently, R (v4.3.2) was used to conduct statistical analyses and visualizations using Phyloseq (v1.46.0) and additional packages such as VennDiagram (Chen and Boutros, 2011), UpsetR (Conway et al., 2017), ggplot2 (Wickham, 2016), gridExtra (Auguie, 2017), tidyverse (Wickham et al., 2019), vegan (Oksanen et al., 2020), ggpubr (Kassambara, 2020), reshape2 (Wickham, 2007), plotly (Sievert, 2020), microbiomeutilities (Lahti and Shetty, 2017), ampvis2 (Andersen et al., 2018), and microbiotaProcess (Xu et al., 2021). In short, a Phyloseq object was used to import all the data (McMurdie and Holmes, 2013). The "alpha" function from the Microbiome package (Lahti and Shetty, 2017) was used to compute alpha diversity. Rarefaction curves of the Shannon bacterial ASVs were computed using the Vegan package. Using the methods in the Phyloseq package, beta diversity was analyzed by Weighted Unifrac Bray-Curti's distance (Lozupone et al., 2011) calculations and plotting and visualization with the Phyloseq package.

Relative abundance of the taxa was determined and agglomerated at the phylum, family, and genus levels using the Phyloseq. Venn diagrams were created and UpsetR packages were used to illustrate the number of ASVs unique and common among different communities (Chen and Boutros, 2011). The core bacterial microbiome of soil samples was calculated based on relative abundance using "Microbiome analyst" 2.0 (Chong et al., 2023). Differential abundance of microbial groups was assessed using DESeq2 (Love et al., 2014), with biomarker characteristics identified based on significant treatment-related changes ($p < 0.05$) and an effect size > ± 1 (log2FoldChange > ± 1). All analyses were considered statistically significant at a p-value of less than or equal to 0.01, except for DESeq2 analysis (Love et al., 2014). Finally, network plots were generated using the Phyloseq package in R (McMurdie and Holmes, 2013), which involved creating an object from the microbiome data, followed by the application of the igraph package (Csardi and Nepusz, 2006) to visualize relationships among taxa based on co-occurrence patterns.

3 Results

3.1 Microbiome richness and diversity

A total of 7,485,810 (7.49 million) and 8,088,548 (8.09 million) raw reads (R1+R2) of 16S rRNA and ITS2, respectively were obtained from soil samples collected from four counties of North Alabama in triplicates. After quality control and trimming using DADA2, we retained 7,204,800 (7.20 million) bacterial and 7,759,796 (7.76 million) fungal high quality sequences. The final unique sequences collected after trimming, dereplicating, filtering chimeric regions, and size selection for bacterial (Table 2) and fungal (Table 3) sequences were presented. Our taxonomic

TABLE 2 Read summary table for the soil samples from 16S rRNA Sequencing.

RN Infestation Level/Groups	County Name	rawseqs (R1+R2)	trimmed_seqs (R1+R2)	chimera_seqs	seqs (after_size_filtration)	final_unique_seqs
ND-A	Jackson	2288000	2201756	84424	813571	12325
LI-B	Lauderdale	1528770	1471846	44899	483396	10344
MI-C	Madison	1611864	1550944	50517	488691	11014
HI-D	Limestone	2057176	1980254	46635	820781	12232

RN, Reniform Nematode; ND, Not-Detected; LI, Low Infestation; MI, Medium Infestation; HI, High Infestation.

TABLE 3 Read summary table for the soil samples from ITS2 Sequencing.

RN Infestation Level/Groups	County Name	rawseqs (R1+R2)	trimmed_seqs (R1+R2)	chimera_seqs	seqs (after_size_filtration)	final_unique_seqs
ND-A	Jackson	2055524	1971520	24407	942064	2446
LI-B	Lauderdale	1909194	1831098	26534	866174	3105
MI-C	Madison	1668612	1601018	43696	731102	1556
HI-D	Limestone	2455218	2356160	52945	1110584	747

RN, Reniform Nematode; ND, Not-Detected; LI, Low Infestation; MI, Medium Infestation; HI, High Infestation.

assignment from the DADA2 pipeline revealed novel and intriguing insights when we compared our samples against the RDP v19 training set for 16S rRNA and UNITE database for ITS data analyses.

3.2 Alpha diversity

Metrics of alpha diversity are employed to assess the richness and evenness of a sample's microbial community at various levels of RN infestation with Kruskal–Wallis test ($p < 0.01$), providing insights into microbial community composition (Lundberg et al., 2020a; Allen and Banfield, 2021). The observed species revealed higher bacterial richness in Group A indicating greater species diversity, while a higher bacterial richness and evenness (Shannon) was identified in Group D with $p < 0.01$ reflecting a more even distribution of species within the microbial community compared to the other groups (Figures 1A). Whereas, the fungal communities in Group A exhibited higher richness and evenness (Shannon index) and higher richness with observed species in Group B with a statistically significance ($p < 0.01$). However, in Group D, the richness for observed species was lower when compared with the Shannon index for richness and evenness (Figures 1B).

Shannon rarefaction curves indicated similar trends in microbial diversity across various levels of RN infestation for both bacterial and fungal communities (Figures 2A, B). Specifically, the bacterial communities in Group D exhibited the highest microbial diversity, with values ranging from 7.34 to 7.67, while Group B showed the lowest diversity, ranging from 6.78 to 7.27 (Figure 2A). In contrast, the fungal communities revealed that Group A had the highest microbial diversity, ranging from 4.28 to 4.85, whereas Group D exhibited the lowest diversity, with values ranging from 3.17 to 3.94 (Figure 2B). Notably, after reaching 30,000 sequences, the Shannon index plateaued at the 97% similarity threshold ($\alpha = 0.03$), indicating that sufficient sequences were obtained to meet the sequencing requirements (Olesen and Simmelsgaard, 2019). The Shannon rarefaction curves for both bacterial and fungal samples (Figures 2A, B) illustrated that the RN infestation curves increased linearly before stabilizing suggesting that the sequencing data was reliable for further investigation (Chao et al., 2014).

3.3 Beta diversity

PERMANOVA analysis of weighted UniFrac distances revealed significant differences ($p < 0.01$) in microbial composition at various levels of RN infestation (Gauthier et al., 2022). The beta diversity or principal coordinate analysis (PCoA) plot, based on weighted UniFrac distances, demonstrated that bacterial groups associated with different levels of RN infestation clustered distinctly from fungal groups (Figures 3A, B). In the beta diversity plot, samples with similar bacterial composition profiles were clustered together, while those with differing profiles were positioned further apart, effectively illustrating the overall bacterial composition. The

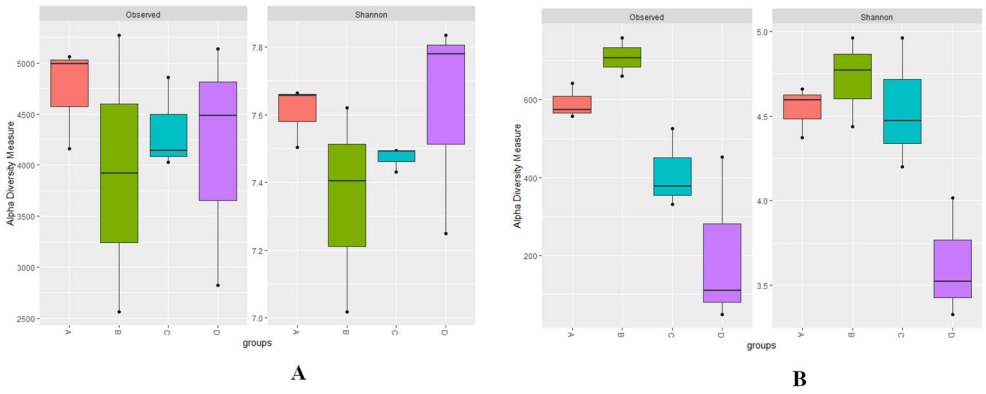


FIGURE 1
(A, B) Boxplots representing the bacterial observed, bacterial Shannon, fungal observed, and fungal Shannon indices at different levels of RN Infestation ($p<0.01$, Kruskal–Wallis test). Group A-RN Not-Detected, Group B-RN Low Infestation, Group C-RN Medium Infestation, Group D-RN High Infestation.

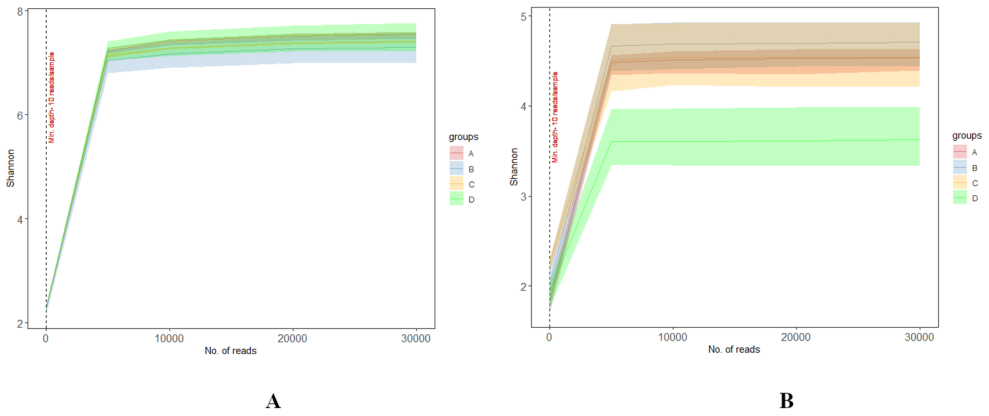


FIGURE 2
(A, B) Rarefaction curves illustrate the Shannon diversity indices of bacterial and fungal communities at different levels of RN Infestation with statistical significance ($p < 0.01$). Group A-RN Not-Detected, Group B- RN Low Infestation, Group C- RN Medium Infestation, Group D- RN High Infestation.

microbial diversity within the fungal Group D clustered and overlapped with groups A and C across various RN infestation levels (Figure 3B). The presence of RN infestation notably influenced the clustering patterns of the samples and their microbial classification (Nielsen et al., 2023).

3.4 Relative abundance

3.4.1 Phylum level

At the phylum level, the bacterial phyla *Actinobacteria, Proteobacteria, Acidobacteria*, and *Planctomycetes* exhibited high relative abundances, followed by *Chloroflexi, Firmicutes, Gemmatimonadetes, Verrucomicrobia*, and *Bacteroidetes* across various levels of RN infestation with a statistical significant difference ($p < 0.01$) (Figure 4A). In all four groups A, B, C, and D-*Actinobacteria, Proteobacteria*, and *Acidobacteria* demonstrated similar patterns of relative abundance. Notably, in Group A, *Planctomycetes* were more abundant than in the other groups, while *Firmicutes* showed higher relative abundance in Group B. In Group C, *Verrucomicrobia* was the most abundant, whereas the highest abundances of *Chloroflexi* and *Bacteroidetes* observed in Group D (Figure 4A). The fungal community composition indicates that *Ascomycota* is the most predominant phylum, followed by *Basidiomycota, Mucoromycota*, and *Rozellomycota* with a significant statistical difference ($p < 0.01$). Specifically, *Ascomycota* was the dominant phyla in Group D, while *Basidiomycota* and *Mucoromycota* were most abundant in Group A. However, *Mucoromycota* was the least abundant phyla across all groups except for Group A, highlighting distinct compositional differences among the groups (Figure 4B).

3.4.2 Family level

At the family level, the bacterial families *Solirubrobacteraceae, Nocardiodaceae, Micromonosporaceae, Acidimicrobiaceae, Acetobacteraceae, Streptomycetaceae*, and *Geodermatophilaceae* were

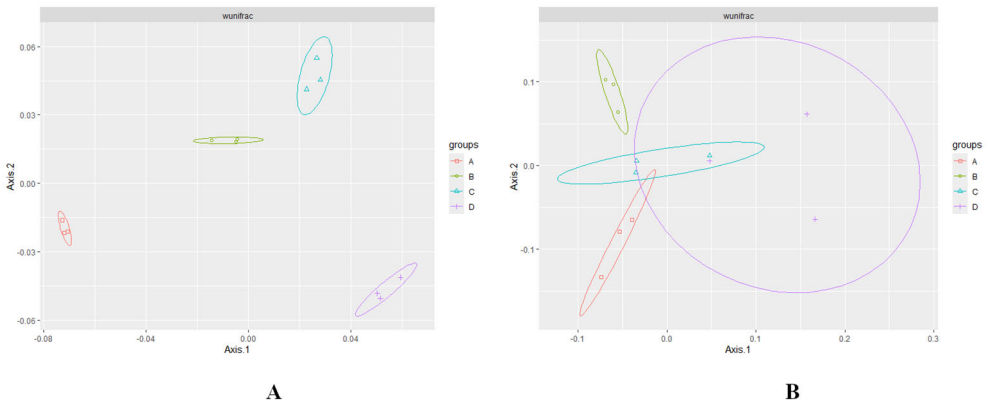


FIGURE 3
(A, B) Principal Coordinate Analysis (PCoA) plot based on Bray-Curtis weighted uniFrac showing the distance in the bacterial and fungal communities at different levels of RN Infestation. Significance was tested using PERMANOVA test ($p < 0.01$). Group A-RN Not-Detected, Group B- RN Low Infestation, Group C- RN Medium Infestation, Group D- RN High Infestation.

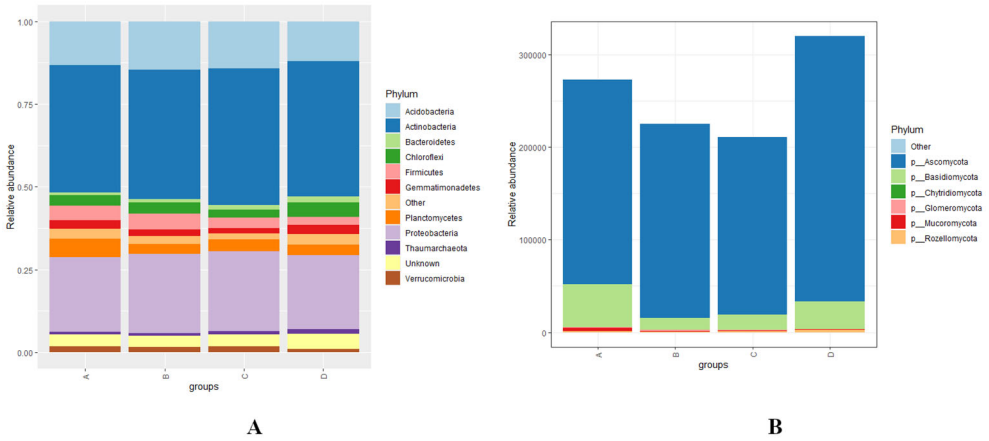


FIGURE 4
(A, B) Distribution and relative abundance of bacterial and fungal phyla at different levels of RN Infestation with statistical significance ($p < 0.01$). Group A-RN Not-Detected, Group B- RN Low Infestation, Group C- RN Medium Infestation, Group D- RN High Infestation.

the most abundant with the statistical significance of $p < 0.01$ across various levels of RN infestation. In Group A, *Solirubrobacteraceae, Cellulomonadaceae*, and *Nocardiodaceae* were observed as the most abundant bacterial families, while *Gaiellaceae* was the least identified, and *Acidimicrobiaceae* was completely absent. In Group B, *Gaiellaceae* emerged as the most abundant family, whereas *Acidimicrobiaceae* was the least abundant. In Group C, *Micromonosporaceae* and *Acidimicrobiaceae* were the most abundant, while *Cellulomonadaceae* and *Gaiellaceae* were the least abundant families. In Group D, *Acetobacteraceae, Streptomycetaceae*, and *Geodermatophilaceae* were the most abundant bacterial families. Remarkably, *Solirubrobacteraceae* was also found to be the least abundant in both groups C and D (Figure 5A). The diversity of fungal families was highlighted by the predominance of *Nectriaceae, Bionectriaceae, Plectosphaerellaceae, Cladosporiaceae, Chaetomiaceae*, and *Ophiocordycipillaceae* across various levels of RN infestation with the statistical significance of $p < 0.01$. In Group A, *Nectriaceae* and *Bionectriaceae* were identified as the most abundant families, *Plectosphaerellaceae* was the least represented, and *Botryosphaeriaceae* and *Pezizomycotina-farm-incertae sedis* were not detected. In Group B, *Plectosphaerellaceae* was the most abundant family and *Trichocomaceae* and *Pezizomycotina-farm-incertae sedis* were not observed. While in group C, *Trichocomaceae* was the most prevalent and *Cladosporiaceae* was the least abundant. Notably, *Bionectriaceae* was the least abundant in both groups B and C. In Group D, *Cladosporiaceae*, *Pezizomycotina-farm-incertae sedis*, *Chaetomiaceae, Botryosphaeriaceae, Ophiocordycipillaceae* were the most abundant and *Bionectriaceae* was the least abundant fungal families identified. Similarly to Group B, *Trichocomaceae* was also not observed in Group D (Figure 5B).

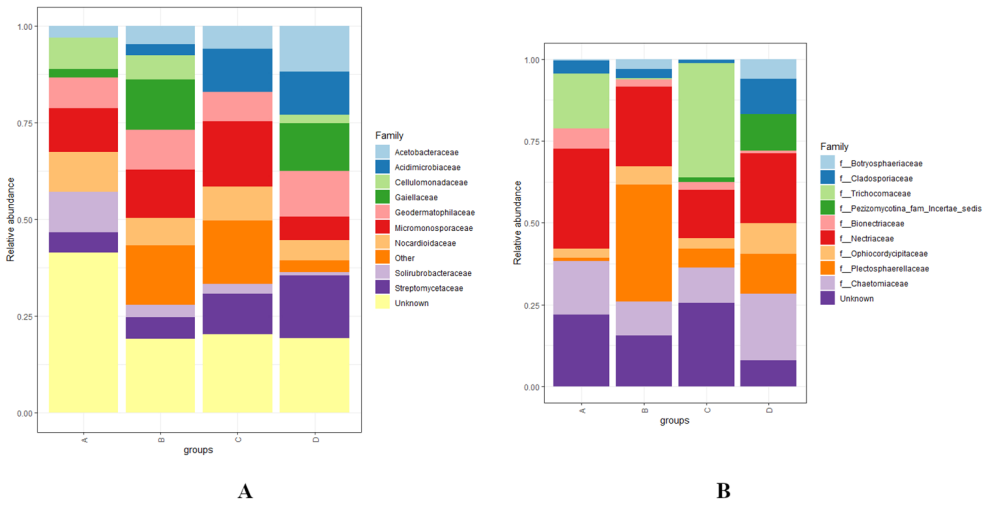


FIGURE 5
(A, B) Distribution and relative abundance of bacterial and fungal communities at the family level across various levels of RN infestation with statistical significance ($p < 0.01$). Group A-RN Not-Detected, Group B- RN Low Infestation, Group C- RN Medium Infestation, Group D- RN High Infestation.

3.4.3 Genus level

At the genus level, *Solirubacter, Nocardioides, Dactylosporangium, Rugosimonospora, Blastococcus, Streptomyces*, and several unclassified genera were identified as the most abundant bacterial genera across different levels of RN infestation with statistical significance, $p < 0.01$. In Group A, *Kribbella, Cellulomonas*, and *Solirubrobacter* were identified as the most abundant, *Rugosimonospora* was found to be the least abundant and *Illumatobacter* was not detected. In Group B, *Gaiella* and *Nocardioides* were identified as the most abundant and *Illumatobacter* was identified as the least abundant genera. In contrast to Group A, *Kribbella* was absent in Group B. In Group C, *Ilumatobacter, Dactylosporangium*, and *Rugosimonospora* were identified as the most abundant while *Gaiella* and *Cellulomonas* were completely absent. In Group D, *Blastococcus* and *Streptomyces* were detected as most abundant. Interestingly, *Solirubacter* was observed as the least abundant genera in groups C and D (Figure 6A). The diversity of fungal genera composition was statistically significant at $p < 0.01$ with predominant genera including *Fusarium, Lectera, Gibellulopsis, Purpureocillum, Cladosporium, Fusarium, Macrophomina* and several unclassified genera across various levels of RN infestation. In Group A, *Fusarium* was identified as the most abundant and *Talaromyces* was the least represented genera. In Group B, *Lectera, Gibellulopsis*, and *Purpureocillum* were identified as the most abundant and *Aspergillus* was not detected. In addition, *Didymela* was not detected in both Group A and B. In Group C, *Talaromyces* and *Aspergillus* were identified as the predominant genera and *Lectera* was identified as the least abundant genera. In Group D, *Cladosporium, Didymella, Fusarium*, and *Macrophomina* were abundant, *Gibellulopsis* was the least abundant and *Talaromyces* was not detected (Figure 6B).

3.5 Venn diagram

To further understand the bacterial and fungal distribution within the microbiota, shared and unique ASVs across different groups under comparison were analyzed using a Venn diagram (Figures 7A, B). In total, 47,893 bacterial and 3,409 fungal ASVs were identified among all groups, with 95 ASVs shared among all bacterial groups and 61 ASVs shared among all fungal groups. For the bacterial ASV's, 12,758, 10,709, 12,153 and 11,360 unique ASVs were identified in Groups A, B, C and D, respectively (Figure 7A). For the fungal ASVs, 663, 887, 480, and 326 unique ASVs were identified in Groups A, B, C and D, respectively. (Figure 7B).

3.6 Differential relative abundance analysis

To investigate variations in the relative abundance of bacterial and fungal genera, we analyzed the dataset using log2 fold change by comparing Group A to Group D. This differential abundance analysis revealed significant changes in the bacterial and fungal microbial communities (Figure 8A). Several bacterial genera exhibited a marked increase ($p < 0.05$; $\log_2FC > 2$) in the abundance of *Streptosporangium, Labrys, Pseudonocardia, Mesorhizobium, Sphingomonas, Arthrobacter, Ilumatbacter*, and *Gaiella* in Group D, compared to Group A. Contrastingly, a significant decrease in the abundance of bacterial genera such as *Burkholderia, Micromonospora, Jatrophihabitans, Gaiella*, and *Mycobacterium* was observed (Figure 8A). The fungal genera exhibited a marked increase in the abundance of *Gibellulopsis, Latorua, Myrothecium, Podospora, Russoella*, and *Bovista* in Group D relative to Group A. Contrastingly, a significant decrease in the abundance of *Tricladium, Phialophora, Humicola, Talaromyces, Nigrospora, Chaetosphaeris, Candida, Myrmecredium, Penicillium*, and *Aspergillus* were observed (Figure 8B).

3.7 Core microbiome

A total of 27 bacterial and 38 fungal genera were identified as part of the core microbiome, across all groups, considering a minimum of 0.1% abundance observed among > 20% of the samples (Figures 9A,

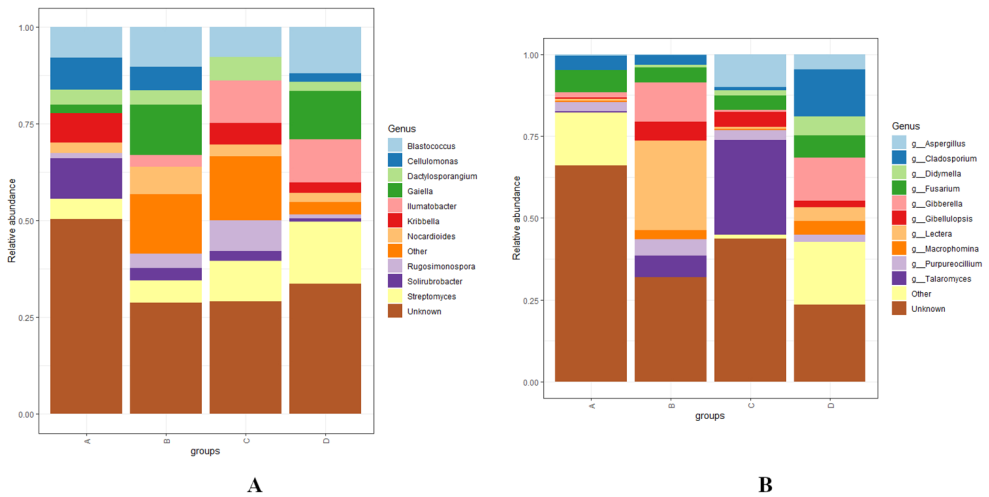


FIGURE 6
(A, B) Distribution and relative abundance of bacterial and fungal genera at different levels of RN Infestation with statistical significance ($p < 0.01$). Group A-RN Not-Detected, Group B- RN Low Infestation, Group C- RN Medium Infestation, Group D- RN High Infestation.

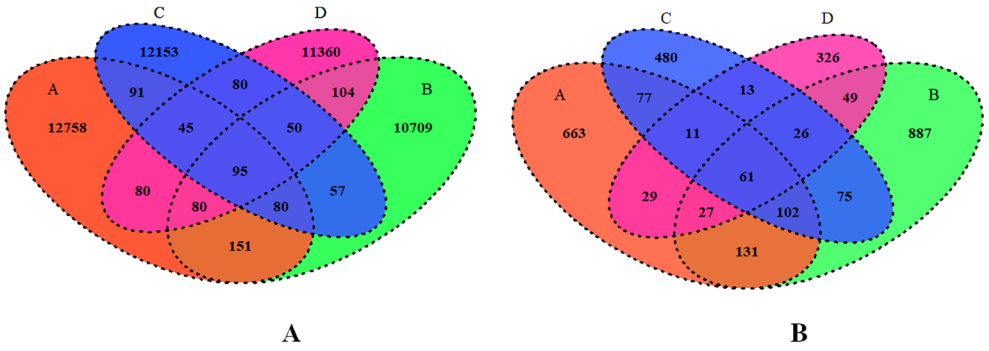


FIGURE 7
(A, B) Venn diagram showing total numbers of shared bacterial and fungal ASVs across various levels of RN Infestation. Group A-RN Not-Detected, Group B- RN Low Infestation, Group C- RN Medium Infestation, Group D- RN High Infestation.

B). Notably, *Gaiella* emerged as the core bacterial genera, exhibiting a prevalence of 100% and a relative abundance of 8%. Additionally, *Conexibacter*, *Bacillus*, *Blastococcus*, and *Streptomyces* were recognized as core bacterial genera, each showing a relative abundance of 0.5% across all groups. Furthermore, *Sphingomonas*, *Mycobacterium*, *Actinoallomurus*, *Dactylosporangium*, *Skermanella*, *Bradyrhizobium*, *Pseudonocardia*, and *Nitrospora* were classified as core bacterial genera, each with a relative abundance of 0.1% across all groups. The remaining bacterial genera displayed an abundance of 0.1% with a prevalence ranging from 70-90% (Figure 9A). In terms of fungal genera, *Fusarium* was recognized as a core member of the microbiome, with a prevalence of 100% and a relative abundance of 1%. *Aspergillus* was also identified as a core fungal genus, with a relative abundance of 0.5% across all groups. Moreover, *Gibberella*, *Cladosporium*, and *Lactera* were categorized as core fungal genera, each exhibiting a relative abundance of 0.1% across all groups. The remaining fungal genera had an abundance of 0.1% with a prevalence range of 30-90% (Figure 9B).

3.8 Network plots of bacterial and fungal phyla

Network plots serve as a powerful visual tool for understanding the relationships among various bacterial and fungal phyla with statistical correlation, $p < 0.01$. In these plots, nodes represent

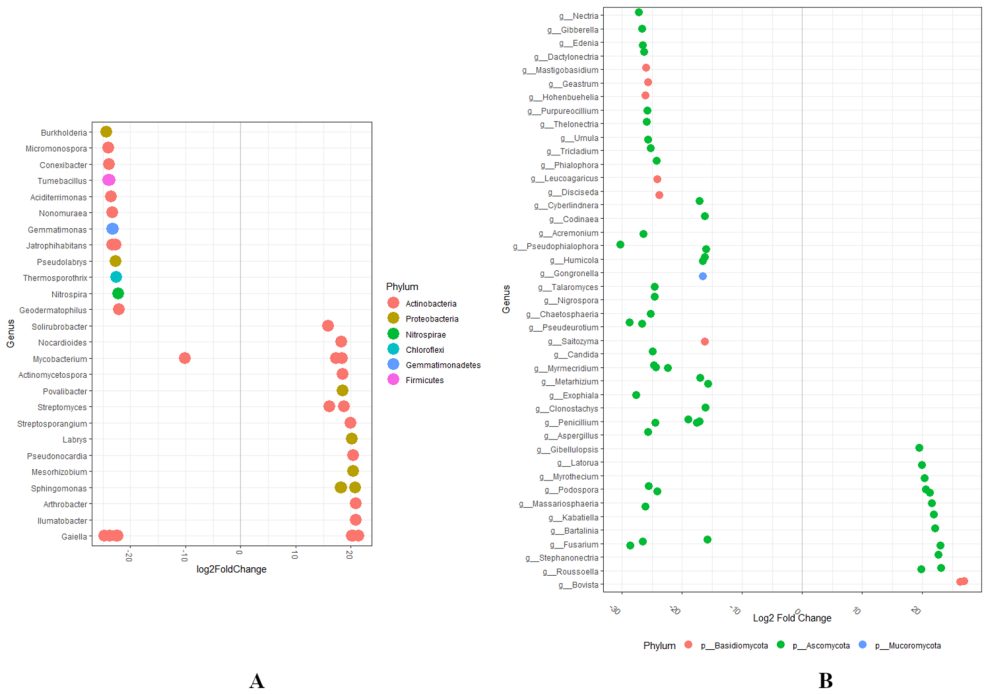


FIGURE 8
(A, B) Differential abundance analysis of bacterial and fungal microbial genera at the phylum level was conducted by comparing Group A (RN Not-Detected) with Group D (RN High Infestation) using the DESeq2 (statistical significance, $p < 0.05$).

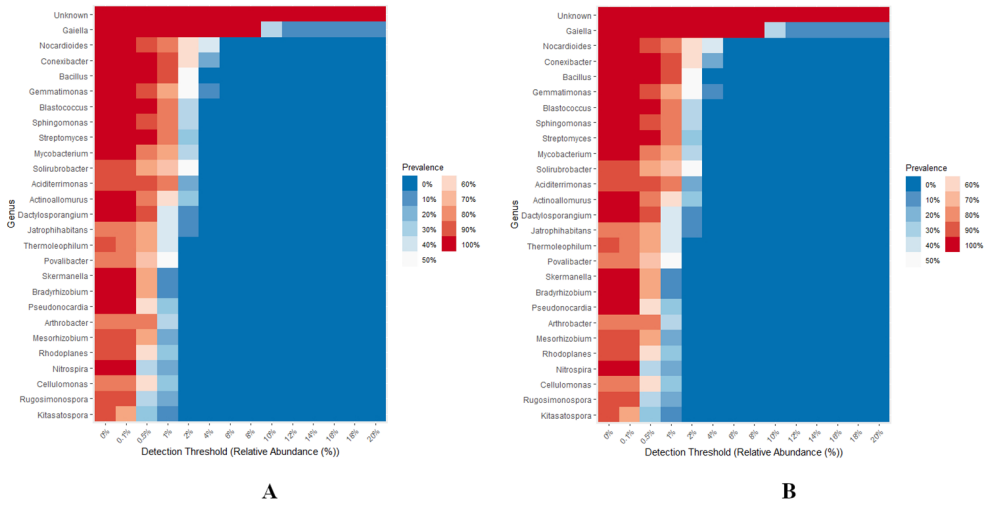


FIGURE 9
(A, B) Heat map showing core bacterial and fungal genera across different levels of RN Infestation. The plot compares the prevalence of genus in samples across varying levels of abundance. Only the genera with minimum prevalence of 0.2 at 0.001 abundance are plotted.

different bacterial phyla, such as *Actinobacteria, Acidobacteria, Bacteroidetes, Firmicutes, Planctomycetes*, and *Proteobacteria* and. fungal phyla such as *Ascomycota, Basidiomycota, Glomeromycota*, and *Mucoromycota*. Edges represent the degree of similarity in taxonomic composition based on shared ASVs (Figures 10A, B). A tight clustering of bacterial nodes was identified among *Actinobacteria, Acidobacteria*, and *Proteobacteria* indicating a high degree of similarity and co-occurrence. This clustering suggests niche sharing, potential ecological interactions and functional roles in the ecosystem (Figure 10A). Conversely, isolated bacterial nodes were found in *Bacteroidetes, Firmicutes*, and *Planctomycetes* with moderate connections indicating distinct ecological dynamics or a specialized function within the community. The nodes of fungal phyla like *Ascomycota* and *Basidiomycota* were closely connected, implying similar ecological roles within the microbial community. Contrastingly, nodes of the *Glomeromycota* and *Mucoromycota* appear to be more isolated, indicating varied functional roles and ecological dynamics (Figure 10B).

4 Discussion

This study advances our understanding of the rhizosphere-associated microbiome to RN infestations by revealing bacterial and fungal richness and evenness shifts across varied infestation levels. Our analysis shows that nematode infestation significantly affects the composition and diversity of rhizospheric microbiota, with notable shifts across infestation levels. Specifically, Group A has the highest abundance of bacterial phyla *Planctomycetes* and fungal phyla *Basidiomycota*, *Mucoromycota*, and *Ascomycota*. Group D

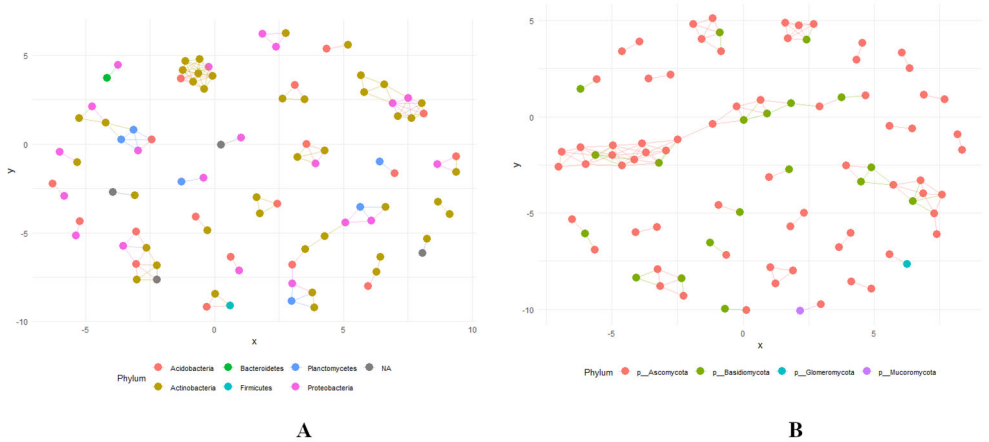


FIGURE 10
(A, B) Network Plot showing the relationship based on Jaccard Distance at phylum level with bacterial and fungal communities across various levels of RN Infestation (Significant ($p < 0.01$) correlation). Group A-RN Not-Detected, Group B- RN Low Infestation, Group C- RN Medium Infestation, Group D- RN High Infestation.

was characterized by the predominance of bacterial phyla *Chloroflexi* and fungal phyla *Ascomycota*. These shifts highlight the resilience and adaptability of the microbiome in response to RN, suggesting a complex interplay between microbial community dynamics and nematode presence. The observed alterations in microbial composition may enhance plant defense mechanisms, as specific microbial taxa can facilitate nutrient acquisition and promote plant growth during stress conditions (Zhang et al., 2020).

The analysis of alpha diversity metrics revealed distinct patterns in microbial community composition corresponding to varying reniform nematode (RN) infestation levels. Notable differences were observed in both bacterial and fungal richness and evenness. Group D showed higher bacterial richness and evenness, while Group A had lower bacterial richness. These results are consistent with those of Lundberg et al. (2020b), that reported increased pest incidence like nematode infestations, can lead to more diverse microbial communities due to improved nutrient availability during stress. Conversely, Group A's fungal communities showed higher richness and evenness, indicating a more stable microbial community associated with healthy soil ecosystems, as reported by Yuan et al. (2020b). Moreover, Group B revealed elevated fungal richness in observed species, suggesting that even moderate nematode infestations can modify root exudation patterns that may subsequently benefit plant health, as Topalovic et al. (2020) indicated. Conversely, in Group D, both fungal richness and evenness were diminished, consistent with previous studies indicating that increased nematode infestation can disrupt the microbial balance, favoring opportunistic species and resulting in a decline in overall microbial diversity, as reported by Zhang et al. (2019).

The Shannon rarefaction curves generated in our study indicate that microbial diversity responds distinctly to varying levels of RN infestation for both bacterial and fungal communities. The observation that the Shannon index plateaued after reaching 30,000 sequences supports the notion that our sequencing efforts sufficiently captured the microbial diversity present in the samples, corroborating similar studies suggested by Olesen and Simmelsgaard (2019) and Chao et al. (2014), which suggest that adequate sampling depth is crucial for reliable diversity assessments. Group D exhibited higher microbial diversity, which aligned with previous reports showed that microbial diversity during nematode infestations (Huang et al., 2022; Zhang et al., 2023). In Group D, the values of the Shannon index showed a higher microbial diversity within bacterial communities suggesting a robust and resilient bacteria that is capable of sustaining functions that are critical for nutrient cycling and plant growth (Wang et al., 2023). Fungal communities showed distinct patterns, with Group A having the highest diversity, suggesting low RN infestation levels may favor a diverse fungal community, potentially enhancing plant health and nutrient uptake (Martínez-García et al., 2023). In contrast, lower fungal diversity indicates that biotic stress negatively impacts the dynamics of fungal communities (Liu et al., 2023).

The PERMANOVA analysis of weighted UniFrac distances ($p < 0.01$) showed differences in the microbial composition across all groups, suggesting that nematode-induced biotic stress is the primary factor driving microbial community shifts. As spatial variation could potentially influence microbial communities, our experimental design ensured that all samples were collected from similar environmental conditions, minimizing spatial variability. In addition, nematode behavior can also cause significant shifts in microbial communities, independent of spatial variation in the sampling environment (Zhang et al., 2019). In the principal coordinate analysis (PCoA), bacterial groups associated with different levels of RN infestation clustered distinctly from fungal groups (Figures 3A, B), a similar pattern reported by Raaijmakers et al., 2009; Zhao et al., 2018; Zhang et al., 2019. Interestingly, an overlap of fungal communities was observed among Groups A, C, and D, suggesting the functional stability of specific fungal taxa despite the fluctuations in biotic stress (Nielsen et al., 2023).

At the phylum level, microbial community composition reveals significant insights into the dynamics of bacterial and fungal populations in response to varying levels of RN infestation. Our results indicate that the bacterial phyla *Actinobacteria, Proteobacteria*, and *Acidobacteria* were highly abundant across all groups, consistent with earlier studies that highlighted their roles in nutrient cycling and plant growth promotion in soil ecosystems (Fierer et al., 2007). Specifically, *Actinobacteria* are known for their capacity to degrade organic matter and contribute to soil health (Jansson and Hofmockel, 2009). Notably, *Planctomycetes* were identified as abundant phyla in Group A, which may indicate their role in nitrogen cycling (Rao and Rao, 2016). Similarly, the higher relative abundance of *Firmicutes* in Group B, *Verrucomicrobia* in Group C, and *Chloroflexi* and *Bacteroidetes* in Group D suggests that these bacteria may play a crucial role in maintaining soil health and nutrient cycling and also associated with the breakdown of complex organic compounds thus enriching the soil (Guan et al., 2018; Ding et al., 2024; Wang et al., 2019). The higher abundance of *Ascomycota* in Group D is important for improving soil health (Lücking et al., 2017). In Group A, the higher abundance of *Basidiomycota* and *Mucoromycota* reflects a potential competitive advantage of these phyla (Zong et al., 2021). A lower abundance of *Mucoromycota* was observed among all groups except Group A, suggesting that the resilience of fungal groups varies with nematode infestation levels (Nielsen et al., 2023).

In Group A, the dominance of *Solirubrobacteraceae* and *Cellulomonadaceae* families was observed, similar to what was reported by Youssef et al., 2015, suggesting their role in cellulose degradation. *Gaiellaceae* was the most abundant family observed in Group B, indicating its adaptability to varied environmental conditions (Vasquez et al., 2019). In contrast, *Acidimicrobiaceae* was the least abundant family in both Group A and B, which may indicate competitive exclusion by more dominant families in less disturbed soils (Sang et al., 2019). In Group C, *Micromonosporaceae* and *Acidimicrobiaceae* were the most abundant families, which can play a crucial role in secondary metabolite production and mitigate stress impacts (Liu et al., 2020). Group D exhibited increased abundance with *Acetobacteraceae*, *Streptomycetaceae*, and *Geodermatophilaceae* families, which are generally associated with nutrient cycling (Meyer et al., 2022). The predominance of

Nectriaceae and *Bionectriaceae* in Group A suggests their vital role in plant health, nutrient mobilization, and soil ecology (Kurtzman et al., 2018). In Group B, the *Plectosphaerellaceae* family was the most abundant, and *Trichocomaceae* was the least abundant, suggesting a potential vulnerability of specific fungal taxa even with mild RN infestation (Zhao et al., 2021). In Group D, the increased abundance of *Cladosporiaceae* and *Chaetomiaceae* suggests the stability of these fungal communities that may affect plant-microbe interactions and overall soil health (Nielsen et al., 2023).

In Group A, the prevalence of *Kribbella*, *Cellulomonas*, and *Solirubacter* and the absence of *Illumatobacter* indicates that specific genera were dominant and are potentially involved in cellulose degradation and organic matter breakdown (Fierer et al., 2007; Liu et al., 2020). In Group B, the relative abundance of *Gaiella* and *Nocardioides* reflects a shift in community dynamics, which can enhance nutrient availability and promote plant growth (Zhang et al., 2018). In Group C, a higher abundance was observed in *Illumatobacter*, *Dactylosporangium*, and *Rugosimonospora*, indicating their resilience and adaptability in response to biotic stress (Pester et al., 2010). A decrease in the abundance of *Gaiella* and *Cellulomonas* suggests that certain nematodes can affect root-associated microbial communities essential for maintaining plant health (Jousset et al., 2017). In Groups C and D, *Solirubacter* was the least abundant genus potentially sensitive to higher RN infestation levels (Zhang et al., 2020). In Group A, *Fusarium* is the most abundant genus that plays a dual role as a pathogen and beneficial organism (Gams et al., 2011). Lower abundance of *Didymella* in Groups A and B may indicate shifts in community structure linked to nematode infestation levels (Grondahl et al., 2021). In Group C, the dominance of *Talaromyces* and *Aspergillus* suggests that specific fungal taxa might thrive in moderate RN infestation, likely due to their saprophytic capabilities and ability to decompose organic matter (Jaklitsch et al., 2016). The abundance of *Cladosporium*, *Didymella*, *Fusarium*, and *Macrophomina* in Group D may support the complex interactions in soil systems. A lower abundance of *Talaromyces* reflects competitive exclusion caused by higher nematode loads (Zhang et al., 2022).

Among the bacterial communities, 95 shared ASVs among all groups indicate a core set of taxa that persists across varying environmental conditions (Shade et al., 2012a). Identifying 61 shared fungal ASVs underscores the potential for specific fungal taxa to adapt and thrive in diverse soil environments (Glassman et al., 2017). The high number of unique ASVs in Group A (12,758) may facilitate niches that support a wider range of bacterial diversity that may be involved in improving nutrient availability and fostering ecological interactions (Lauber et al., 2009). A lower number of unique ASVs (10,709) were observed in Group B compared to Group A, reflecting a substantial diversity that may stimulate competitive interactions under mild RN infestation (Santos et al., 2018). In Group C, 12,153 unique ASVs identified belonged to diverse bacterial communities with specific functional roles in soil and plant health (Friedman and Alm, 2012). In Group D, relatively lower bacterial communities (11,360 unique ASVs) identified were possibly due to the microbial and pest competition for the available nutrients (Wagg et al., 2014). A higher diversity of fungal ASVs was identified in Group A (663) compared to Group D (326), suggesting that specific fungal communities may be more resilient to biotic stress, highlighting the interplay between fungal diversity and plant health (Zhao et al., 2021).

By suppressing plant defenses, nematodes may inadvertently alter the composition of microbial communities in the rhizosphere. This can lead to an increase in stress-resistant bacterial taxa better adapted to the modified environment. Moreover, the weakened plant defense system can also facilitate the colonization and proliferation of secondary pathogens, including bacteria, which may further exploit the compromised plant defenses (Shade et al., 2012b; McGuire et al., 2017). The increased abundance of genera such as *Streptosporangium, Labrys, Pseudonocardia, Mesorhizobium, Sphingomonas*, and *Arthrobacter* in Group D indicates a shift in stress-resistant taxa. Various ecological interactions and environmental factors influence the shift in specific microbial populations such as *Burkholderia, Micromonospora, Jatrophihabitans, Gaiella*, and *Mycobacterium* to selective pressures from nematode infestations. The change in such bacterial genera in Group D may be attributed to the selective pressure exerted by higher levels of RN infestation. *Burkholderia* species, particularly *B. seminalis*, have shown potential as biocontrol agents against nematodes like *Meloidogyne enterolobii*. Studies have demonstrated that specific concentrations of *B. seminalis* can exhibit ovicidal activity, reducing nematode egg viability and thus controlling nematode populations (Moreira et al., 2024). *Burkholderia* is highly attractive to certain nematodes, such as *M. incognita*, which can increase nematode aggregation around these bacteria. This attraction can influence the dynamics of nematode populations and their interactions with other microbial communities (Tahseen and Clark, 2014). Fungal genera such as *Gibellulopsis, Latorua, Myrothecium, Podospora, Russoella*, and *Bovista* employ a variety of mechanisms to suppress nematode infestations. These fungi are part of a broader group known as nematophagous fungi, which are recognized for their ability to control nematode populations through diverse strategies such as mechanical trapping, endoparasitism, systemic resistance, enzymatic degradation, and toxic metabolite production (Noweer, 2020). Interestingly, a significant increase in the abundance of these fungal genera was observed in Group D. A significant decrease in the fungal genera such as *Tricladium, Phialophora, Nigrospora*, and *Candida* was observed. The presence of some nematophagous fungi can potentially reduce the prevalence of non-nematophagous genera like *Tricladium* and *Phialophora* (Mo et al., 2023).

The presence of nematodes in the soil leads to increased alkaline phosphomonoesterase (ALP) activity, directly linked to higher phosphorus availability. This enhanced nutrient cycling provides a competitive edge to bacteria like *Gaiella*, that can efficiently utilize the available phosphorus (Zheng et al., 2022). The core bacterial genus *Gaiella* emerged as a dominant genus, exhibiting a prevalence of 100% and a relative abundance of 8%, indicating the crucial role in phosphorus recycling. The specific mechanism by which

Conexibacter, Bacillus, Blastococcus, and *Streptomyces* bacteria outcompete other microbial communities under nematode infestation involves a combination of biochemical and ecological strategies. The presence of these bacteria with a relative abundance of 0.5% suggests their possible role in survival and proliferation in the rhizosphere by competing against the nematodes. Nematode-induced nutrient cycling significantly impacts the selective advantage of bacteria such as *Sphingomonas, Mycobacterium*, and *Actinoallomurus* in mixed microbial communities, as identified in Group D, with a relative abundance of 0.1%. Through predation and feeding activities, nematodes influence the availability of nutrients like nitrogen and phosphorus, affecting bacterial community dynamics and competitive interactions. This process can enhance the growth and activity of specific bacterial taxa, providing them with a competitive edge in nutrient-limited environments (Zheng et al., 2022). In Group D, *Fusarium* was observed with a 100% prevalence and a relative abundance of 1%. This dominance may be due to the complex relationship between *Fusarium* and nematodes, including antagonistic and synergistic interactions. These interactions can vary based on environmental conditions and the specific species involved (Siddiqui and Aziz, 2024). In Group D, *Aspergillus*, *Gibberella, Cladosporium*, and *Lactera* were identified with a relative abundance of 0.1%. These fungi can play various roles, from being parasitic to nematodes to acting as part of a broader soil microbiome that influences nematode behavior and survival.

The presence of nematodes alters the soil environment, affecting the bacterial community structure and promoting the clustering of specific bacterial taxa that can thrive under these conditions. We observed tight clustering among bacterial nodes, particularly in *Actinobacteria, Acidobacteria*, and *Proteobacteria* in the core microbiome. This clustering is likely influenced by the competitive exclusion of less adapted bacterial clades and the selective pressures exerted by the nematodes and the altered soil environment (Yergaliyev et al., 2020). The isolation of nodes within *Bacteroidetes, Firmicutes*, and *Planctomycetes* under plant parasitic nematode infestation is driven by complex interactions between the nematodes, the plant host, and the microbial communities in the rhizosphere. These interactions are influenced by the nematode's life cycle, the plant's response to infestation, and the environmental conditions in the soil (Yergaliyev et al., 2020). Forming closer tier networks within *Ascomycota* and *Basidiomycota* in response to RN levels provides significant evolutionary advantages. These fungi have evolved mechanisms that enhance their survival and ecological roles by forming intricate networks optimized for resource acquisition, defense, and symbiosis with host plants. Such networks are essential in environments with higher nematode levels, as they help mitigate the damage caused by these pests (Kitagami and Matsuda, 2024). Conversely, the formation of isolated networks with *Glomeromycota* phylum, particularly arbuscular mycorrhizal fungi (AMF), such as species from the genus *Glomus*, under RN infestation can be attributed primarily to their potential role in enhancing plant resistance and growth. These fungi possibly establish symbiotic relationships with plant roots, thereby improving nutrient uptake and serve as a biological control against nematodes (Chaerani and Ginting, 2023).

4.1 Potential pitfalls associated with soil microbial profiling studies

Conducting soil microbial analysis presents several potential pitfalls researchers must navigate to ensure accurate and reliable results. One of the significant challenges in such studies is the inherent heterogeneity of soil, which complicates the sampling process. Soil is a dynamic entity with varying microbial populations, and sampling must be statistically sound to capture this diversity accurately. Additionally, the physicochemical properties of soil, such as pH, and organic content can significantly influence microbial community composition and activity, necessitating careful consideration and control in experimental designs. The rapid changes in microbial populations during sample handling and storage also necessitate prompt transfer to laboratories in order to prevent alterations in microbial activity. Also, there is difficulty in estimating the concentration and activity of mixed microbial populations due to their heterogeneous nature and varying metabolic rates. Traditional methods like fluorescence and spectrophotometry have limitations, and microscopy is often recommended for more accurate measurements. Moreover, integrating molecular techniques in soil microbial analysis while offering advanced insights requires careful interpretation to avoid misrepresenting microbial diversity and function. Furthermore, the lack of soil-specific reference databases for metagenomic classifiers poses a challenge in accurately profiling soil microbiomes. Custom databases, optimized classifiers with improved accuracy in taxonomic classification, and tailored bioinformatic pipelines are required. Lastly, sharing data and establishing standard guidelines are crucial for reproducibility and meta-analyses, which can enhance the understanding of soil microbial communities and their ecological roles. The experimental design of this study was structured to address the potential pitfalls by adhering to the Alabama Cooperative Extension System's protocols.

5 Conclusion

The study explores the relationships between reniform nematode (RN) infestation and the rhizosphere microbiome dynamics in cotton soils. It finds that RN infestation affects the diversity and composition of microbial communities, which in turn enhances plant resistance to biotic stress. These microbial shifts also impact vital biogeochemical cycles important for soil fertility. Furthermore, the research delineates specific bacterial and fungal taxa associated with RN infestation, indicating potential approaches for biological control and soil management. Our findings underscore the importance of comprehending plant-microbe-

nematode interactions to formulate integrated pest management strategies that promote sustainable cotton production.

6 Future directions

Designing individual and integrated experiments to understand tripartite interactions among plant-nematode-soil microbiomes is critical during Reniform nematode infestation. Growing and maintaining specific bacterial or fungal pure cultures identified during RN infestation will improve our understanding of these unique microbial species' functional roles. This knowledge will facilitate the exploration of associated plant defense mechanisms, potentially leading to the development of targeted biological control strategies. Also, investigating the interactions between nematodes, rhizosphere microbiomes, and different cotton genotypes using multi-omic approaches could enhance our understanding of metabolite degradation, nutrient availability in soil, host-parasite competition, and selective pressures exerted on microbial populations during nematode infection. To further strengthen our knowledge, pot culture studies under controlled conditions with different genotypes play a crucial role in examining microbial shifts during RN infection to comprehend the link between microbial dynamics and plant resistance. Longitudinal studies assessing the impact of various nematode management practices on microbial community composition and soil health are essential. Applying these approaches to other plant-nematode systems will support our findings. This will help us understand broader ecological effects and promote sustainable farming practices.

Data availability statement

The datasets generated for this study can be found in the NCBI Sequence Reads Archive (SRA) with the accession numbers SAMN45928254 - SAMN45928265 for 16S rRNA and SAMN45929394 - SAMN45929405 for ITS, under the BioProject, PRJNA1201180.

Ethics statement

The manuscript presents research on animals that do not require ethical approval for their study.

Author contributions

SK: Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft. VA: Investigation, Methodology, Writing – review & editing. ST: Investigation, Methodology, Writing – review & editing. MJ: Resources, Writing – review & editing. KL: Resources, Writing – review & editing. LN: Resources, Writing – review & editing. AT: Writing – review & editing. LW: Resources, Writing – review & editing. VS: Conceptualization, Funding acquisition, Project administration, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work is supported by the Capacity Building Grant (#2020-38821-31103) and Evans Allen Grant (#7005717/ALAX - 011-1223EA), from the U.S. Department of Agriculture's National Institute of Food and Agriculture. These funds have been utilized in the study's design, data collection, analysis, interpretation, writing, and manuscript submission.

Acknowledgments

We would like to thank Dr. Govind C. Sharma from Alabama A&M University for his invaluable assistance in conducting this research.

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

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.2025.1521579/full#supplementary-material

SUPPLEMENTARY FIGURE 1
North Alabama map showing levels of Reniform Nematode (RN) Infestation in four selected counties.

References

Abarenkov, K., Tedersoo, L., Nilsson, R. H., Veldre, V., Paap, T., Zirk, A., et al. (2023). UNITE QIIME release for fungal ITS sequencing, with improved taxonomic assignment, analysis, and visualization. *Fungal Diversity* 114, 319–324. doi: 10.1007/s13225-023-00448-z

Alabama Cooperative Extension System (2020). *Soil Survey and Soil Types of North Alabama* (Auburn, AL: Alabama A&M and Auburn University).

Allen, E. E., and Banfield, J. F. (2021). Community genomics in microbial ecology. *Nat. Rev. Microbiol.* 19, 233–244. doi: 10.1038/s41579-020-00456-1

Andersen, K. S., Kirkegaard, R. H., Karst, S. M., and Albertsen, M. (2018). ampvis2: An R package to analyse and visualize 16S rRNA amplicon data. *BioRxiv*, 299537. doi: 10.1101/299537

Auguie, B. (2017). *gridExtra: Miscellaneous functions for "grid" graphics. R package version 2.3*. Vienna, Austria: CRAN. doi: 10.48550/arXiv.cond-mat/0609440

Bais, H. P., Park, S. W., and Vivanco, J. M. (2006). Rhizosphere interactions and plant health. *Annu. Rev. Phytopathol.* 44, 121–152. doi: 10.1146/annurev.phyto.44.070505.143416

Bhattacharyya, P. N., and Jha, D. K. (2012). Soil health and sustainable agriculture: The role of beneficial microbes. *Soil Sci.* 177, 535–543. doi: 10.1097/SS.0b013e31826c6f3e

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Bushnell, B., Rood, J., and Singer, E. (2017). BBMerge – Accurate paired shotgun read merging via overlap. *PloS One* 12, e0185056. doi: 10.1371/journal.pone.0185056

Cai, M., Liu, S., and Xu, X. (2023). Interactions between root-knot nematodes and soil microorganisms: A review. *Soil Biol. Biochem.* 182, 108487. doi: 10.1016/j.soilbio.2023.108487

Callahan, B. J., McMurdie, P. J., and Holmes, S. P. (2016a). DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* 13, 581–583. doi: 10.1038/nmeth.3869

Callahan, B. J., Sankaran, K., Fukuyama, J. A., McMurdie, P. J., and Holmes, S. P. (2016b). Bioconductor workflow for microbiome data analysis: from raw reads to community analyses. *F1000Research* 5, 1492. doi: 10.12688/f1000research

Chaerani, C., and Ginting, R. C. B. (2023). Response of soybean and tomato plants under dual inoculation with Glomus sp. and root-knot nematode Meloidogyne incognita. *IOP Conference Series: Earth and Environmental Science* (Vol. 1271, No. 1, p. 012032). Bristol, UK: IOP Publishing. doi: 10.1088/1755-1315/1271/1/012032

Chao, A., Ma, A. K., and Hwang, W. H. (2014). A new method for estimating the number of shared species in two communities. *Ecol. Indic.* 36, 246–256. doi: 10.1016/j.ecolind.2013.06.013

Chen, H., and Boutros, P. C. (2011). VennDiagram: a package for the generation of highly customizable Venn and Euler diagrams in R. *BMC Bioinf.* 12, 1–7. doi: 10.1186/1471-2105-12-35

Chen, Z., Liu, J., and Li, Y. (2022). The impact of plant-parasitic nematodes on soil microbial communities: A review. *Soil Biol. Biochem.* 165, 108507. doi: 10.1016/j.soilbio.2022.108507

Chong, J., Liu, P., Zhou, G., and Xia, J. (2023). MicrobiomeAnalyst 2.0: A comprehensive platform for the analysis of microbiome data. *Nat. Protoc.* 18, 1–23. doi: 10.1038/s41596-022-00789-5

Celleti, L., and Potter, R. (2006). *Soil Sampling and Collection Procedures*. Auburn, AL: Alabama Cooperative Extension System.

Cole, J. R., Wang, Q., Cardenas, E., Fish, J., Chai, B., Farris, R. J., et al. (2014). Ribosomal Database Project: Data and tools for high-throughput rRNA analysis. *Nucleic Acids Res.* 42, D633–D642. doi: 10.1093/nar/gkt1244

Conway, J. R., Lex, A., and Gehlenborg, N. (2017). UpSetR: An R package for the visualization of intersecting sets and their properties. *Bioinformatics* 33, 2938–2940. doi: 10.1093/bioinformatics/btx364

Csardi, G., and Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal Complex Syst.* 1695, 1-9. doi: 10.48550/arXiv.cond-mat/0609440

de la Fuente, V., Rufo, L., and Iglesias-López, M. T. (2021). Impacts of sample handling and storage conditions on archiving physiologically active soil microbial communities. *FEMS Microbiol. Lett.* 371, fnae044. doi: 10.1093/femsle/fnae044

De Vries, F. T., and Shade, A. (2013). Controls on soil microbial community stability under climate change. *Ecol. Lett.* 16, 309–320. doi: 10.3389/fmicb.2013.00265

Ding, Y., Gao, X., Shu, D., Siddique, K. H. M., Song, X., Wu, P., et al. (2024). Enhancing Soil Health and Nutrient Cycling Through Soil Amendments: Improving the Synergy of Bacteria and Fungi (Rochester, NY: SSRN). Available at: https://ssrn.com/abstract=4681409 (Accessed July 18, 2024).

Eisenback, J. D., and Triantaphyllou, A. C. (1991). "Nematode parasites of plants," in *Nematology: Principles and Practices* (Dordrecht: Springer), p. 191-274.

Fierer, N., Jackson, R. B., Caporaso, J. G., Lauber, C. L., Zhou, J., and Knight, R. (2007). The influence of soil microbial community composition on the diversity of soil fungi. *Soil Biol. Biochem.* 39, 1396–1407. doi: 10.1016/j.soilbio.2006.12.006

Friedman, J., and Alm, E. J. (2012). Inferring correlation networks from genomic survey data. *PloS Comput. Biol.* 8, e1002687. doi: 10.1371/journal.pcbi.1002687

Gams, W., Meyer, W., and Seifert, K. A. (2011). "The genera of hyphomycetes," in *Fungal Diversity Research Series*, vol. 1. (Dordrecht: Springer). doi: 10.1007/978-94-007-0701-8_1

Garbeva, P., van Veen, J. A., and van Elsas, J. D. (2004). Microbial diversity in soil: The quantity and quality of the microbial community are influenced by soil type and land use. *FEMS Microbiol. Ecol.* 48, 1–10. doi: 10.1016/j.femsec.2004.02.001

Garcia-Sanchez, M., Perez-Hernandez, V., Hernandez-Guzman, M., Valenzuela-Encinas, C., Alcantara-Hernandez, R., and Estrada-Alvarado, I. (2020). An improved method for extraction of microbial DNA from alkaline-saline soil. *FEMS Microbiol. Lett.* 371, 1–10. doi: 10.1093/femsle/fnae044

Gauthier, J. C., Sweeney, R. A., and Reeve, J. R. (2022). Distinct shifts in microbial community composition under biotic stress in agricultural systems. *Environ. Microbiol. Rep.* 14, 203–214. doi: 10.1111/1758-2229.13056

Glassman, S. I., Wang, I. J., Bruns, T. D., Taylor, J. W., Peay, K. G., and Vilgalys, R. (2017). Fungal community responses to biotic stressors: Evidence from soil and plant studies. *Fungal Ecol.* 29, 38–47. doi: 10.1016/j.funeco.2017.06.007

Gomez, E. J., Ferris, H., and Lee, J. (2019). Soil microbial communities and their functions in relation to nematodes. *Appl. Soil Ecol.* 135, 165–173. doi: 10.1016/j.apsoil.2018.11.015

Grondahl, S., Pettersson, M., and Bjoerklund, M. (2021). The impact of nematodes on soil fungal communities: A review. *Soil Biol. Biochem.* 153, 108077. doi: 10.1016/j.soilbio.2021.108077

Guan, Y., Xie, L., and Zhang, R. (2018). Impacts of soil nematodes on the diversity and composition of soil microbial communities in the rhizosphere of cucumber. *Sci. Rep.* 8, 3142. doi: 10.1038/s41598-018-21473-6

Hassan, S. E., and Abo-Elyousr, K. A. (2019). Biological control of plant-parasitic nematodes by microorganisms: An overview. *Biocontrol Sci. Technol.* 29, 499–512. doi: 10.1080/09583157.2019.1572719

Huang, Q., Xu, J., and Chen, W. (2022). Microbial diversity and community structure in relation to nematode infestation in the rhizosphere. *Soil Biol. Biochem.* 176, 108943. doi: 10.1016/j.soilbio.2022.108943

Illumina (2019). *NextSeq 2000 System User Guide*. San Diego, CA: Illumina.

Invitrogen (2016). Qubit™ 1X dsDNA Broad Range Assay Kit User Guide (Carlsbad, CA: Thermo Fisher Scientific). Available at: https://www.thermofisher.com (Accessed November 3, 2023).

Jaklitsch, W. M., Samson, R. A., Frisvad, J. C., Seifert, K. A., and Varga, J. (2016). Aspergillus, Penicillium and Talaromyces: a revised classification. *Fungal Diversity* 86, 5–22. doi: 10.1007/s13225-016-0377-y

Jansson, J. K., and Hofmockel, K. S. (2009). Soil microbial communities and climate change. *Nat. Rev. Microbiol.* 7, 423–431. doi: 10.1038/nrmicro2114

Jousset, A., Becker, J., Chatterjee, S., Karlovsky, P., Scheu, S., and Eisenhauer, N. (2017). Interspecific competition alters the rhizosphere microbiome of plants. *Mol. Ecol.* 26, 6153–6162. doi: 10.1111/mec.14360

Kassambara, A. (2020). *ggpubr: "ggplot2" Based Publication Ready Plots. R package version 0.4.0. CRAN*. doi: 10.32614/CRAN.R-package.ggpubr

Kitagami, Y., and Matsuda, Y. (2024). Forest types matter for the community and co-occurrence network patterns of soil bacteria, fungi, and nematodes. *Pedobiologia* 107, 151004. doi: 10.1016/j.pedobi.2024.151004

Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K., and Schloss, P. D. (2013). Development of a dual-indexed sequencing strategy and curation of the Fungal Internal Transcribed Spacer Gene Database. *Appl. Environ. Microbiol.* 79, 6220–6224. doi: 10.1128/AEM.01043-13

Kurtzman, C. P., Robnett, C. J., and Basehoar, J. (2018). *Fungal Ecology: Principles and Mechanisms* (San Diego, CA: Academic Press), 173–204. doi: 10.1016/B978-0-12-811033-8.00007-9

Lahti, L., and Shetty, S. (2017). microbiome R package. Available online at: https://microbiome.github.io (Accessed March 13, 2024).

Latz, M. A., Eisenhauer, N., Rall, B. C., Scheu, S., Jousset, A., and Thakur, M. P. (2021). Utilizing Phyloseq for analysis of microbial community data. *J. Comput. Biol.* 28, 256–271. doi: 10.1089/cmb.2020.0456

Lauber, C. L., Hamady, M., Knight, R., and Fierer, N. (2009). Toward a comprehensive community phylogeny of bacteria: A comparison of 16S rRNA gene and shotgun metagenomic data. *Environ. Microbiol.* 11, 529–541. doi: 10.1111/j.1462-2920.2008.01701.x

Li, D., Liu, C. M., Luo, R., Sadakane, K., and Lam, T. W. (2021). MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31, 1674–1676. doi: 10.1093/bioinformatics/btv033

Li, X., Zhang, X., and Zhang, S. (2022). Microbiome-mediated control of plant-parasitic nematodes: Mechanisms and perspectives. *Environ. Microbiol. Rep.* 14, 289–299. doi: 10.1111/1758-2229.13000

Li, X., Zhang, Y., Chen, L., Wang, J., and Liu, Q. (2023). Soil sample transportation and storage methods for maintaining microbial integrity. *J. Soil Sci.* 59, 112–125. doi: 10.1016/j.soil.2023.01.002

Liu, W., Chen, Y., and Liu, H. (2020). Micromonospora: A potential biocontrol agent against phytopathogenic fungi. *Microbial Ecol.* 80, 569–583. doi: 10.1007/s00248-020-01592-3

Liu, X., Wang, X., and Liu, Y. (2017). Nematode-induced changes in microbial community structure in soil. *Appl. Soil Ecol.* 118, 78–85. doi: 10.1016/j.apsoil.2017.03.015

Liu, Y., Zhang, X., and He, Y. (2023). Impacts of root-knot nematodes on soil bacterial diversity and community structure. *Appl. Soil Ecol.* 174, 104434. doi: 10.1016/j.apsoil.2022.104434

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

Lozupone, C., Lladser, M. E., Knights, D., Stombaugh, J., and Knight, R. (2011). UniFrac: an effective distance metric for microbial community comparison. *ISME J.* 5, 169–172. doi: 10.1038/ismej.2010.133

Lücking, R., Friedl, T., and Rojas, C. (2017). Fungal biodiversity and its ecological significance. *Fungal Diversity* 87, 5–23. doi: 10.1007/s13225-017-0366-0

Lugtenberg, B., and Kamilova, F. (2009). Plant-growth-promoting rhizobacteria. *Annu. Rev. Microbiol.* 63, 541–556. doi: 10.1146/annurev.micro.62.081307.162918

Lundberg, D. S., Lebeis, S. L., Garrity, S. J., and Keller, J. (2020a). Defining the core microbiome of a plant. *Nat. Microbiol.* 5, 292–303. doi: 10.1038/s41564-019-0574-9

Lundberg, D. S., Yourstone, S. M., Mieczkowski, P., Jones, C. D., Dangl, J. L., Gehring, J., et al. (2020b). Microbial community response to increasing herbivore pressure in a rhizosphere model system. *Front. Microbiol.* 11. doi: 10.3389/fmicb.2020.01121

Luo, Y., Hu, J., and Zhang, W. (2022). Soil health indicators in relation to microbial communities: Insights from nematode interactions. *Soil Ecol. Lett.* 4, 15–25. doi: 10.1007/s42832-021-0100-5

Mardis, E. R. (2008). Next-generation DNA sequencing methods. *Annu. Rev. Analytical Chem.* 1, 387–404. doi: 10.1146/annurev.genom.9.081307.164359

Martínez-García, L. B., Ramírez, M. S., and Pérez, D. R. (2023). Functional roles of Ascomycota in nutrient cycling and pathogen suppression in crop rhizospheres. *Front. Microbiol.* 14. doi: 10.3389/fmicb.2023.1132429

McGuire, K. L., Fierer, N., Bateman, C., Treseder, K. K., and Turner, B. L. (2017). Microbial community response to biotic stress: The role of microbial interactions in the rhizosphere. *Ecol. Appl.* 27, 1842–1853. doi: 10.1002/eap.158

McGuire, K. L., and Triplett, E. W. (2009). Microbial communities in the rhizosphere of crop plants. *Nat. Rev. Microbiol.* 7, 641–651. doi: 10.1038/nrmicro2199

McMurdie, P. J., and Holmes, S. (2013). phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PloS One* 8, e61217. doi: 10.1371/journal.pone.0061217

Meyer, K. M., Nason, J. D., and DiMarco, J. (2022). The roles of bacterial families in soil health and nutrient cycling: A review. *Soil Biol. Biochem.* 164, 108454. doi: 10.1016/j.soilbio.2022.108454

Mo, X., Zhang, Y., Li, W., Chen, J., and Wang, H. (2023). Impact of nematophagous fungi on the prevalence of non-nematophagous genera in soil ecosystems. *J. Fungal Ecol.* 12, 45–58. doi: 10.1016/j.funeco.2023.01.001

Moreira, I. F., Marques, M. L., da, S., Santos, T. T. M. D., Rangel, W. de M., Brioso, P. S. T., et al. (2024). Biocontrole de Burkholderia seminalis sobre Meloidogyne enterolobii. *Contribuciones Las Cienc. Sociales.* 17 (3), 48-60. doi: 10.55905/revconv.17n.3-048

Naylor, D., Coleman-Derr, D., and McKenzie, V. (2021). Soil microbiomes and their implications for plant health. *Nat. Rev. Microbiol.* 19, 371–384. doi: 10.1038/s41579-021-00513-0

Naylor, D., and Gurevitch, J. (2021). Rhizosphere microbial communities and their role in the suppression of nematode populations. *Microbial Ecol.* 82, 391–407. doi: 10.1007/s00248-021-01709-8

Nielsen, K. A., Roberts, J. T., and Green, S. R. (2023). Microbial clustering in response to plant-parasitic nematodes: Insights from beta diversity metrics. *Microbial Ecol.* 86, 104–115. doi: 10.1007/s00248-022-01894-1

Noweer, E. M. A. (2020). doi: 10.1007/978-3-030-33161-0_12

Nyaku, S. T., Kantety, R. V., Cebert, E., Lawrence, K. S., Honger, J. O., and Sharma, G. C. (2016). Principal component analysis and molecular characterization of reniform nematode populations in Alabama. *Plant Pathol. J.* 32, 123. doi: 10.5423/PPJ.OA.09.2015.0194

Nyaku, S. T., Kantety, R. V., Lawrence, K. S., van Santen, E., and Sharma, G. C. (2013a). Canonical discriminant analysis of rotylenchulus reniformis in Alabama. *Nematropica* 43, 171–180. doi: 10.18781/NEMATROPICA.V43I2.82705

Nyaku, S. T., Karapareddy, S., Cebert, E., Lawrence, K., Eleblu, J. S., Sharma, G. C., et al. (2023). Two Intra-Individual ITS1 rDNA Sequence Variants Identified in the Female and Male Rotylenchulus reniformis Populations of Alabama. *Plants* 13, 5. doi: 10.3390/plants13010005

Nyaku, S. T., Sripathi, V. R., Kantety, R. V., Gu, Y. Q., Lawrence, K., and Sharma, G. C. (2013b). Characterization of the two intra-individual sequence variants in the 18S rRNA gene in the plant parasitic nematode, Rotylenchulus reniformis. *PloS One* 8, e60891. doi: 10.1371/journal.pone.0060891

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., et al. (2020). *vegan: Community Ecology Package. R package version 2.5-7*. doi: 10.32614/CRAN.R-package.vegan

Olesen, J. E., and Simmelsgaard, S. E. (2019). Importance of sequencing depth for the assessment of microbial community structure in the rhizosphere. *FEMS Microbiol. Ecol.* 95, fiz014. doi: 10.1093/femsec/fiz014

Parker, T. H., Derryberry, E. P., and Derryberry, G. E. (2016). A flexible software pipeline for the analysis of 16S rRNA gene sequences. *F1000Research* 5, 1587. doi: 10.12688/f1000research.9214.1

Patel, A., Smith, B., Johnson, C., Lee, D., and Kumar, E. (2024). Nematode-induced changes in microbial diversity promote plant growth by enhancing phosphorus cycling under nutrient-limited conditions. *J. Soil Biol.* 45, 123–135. doi: 10.1016/j.soilbio.2024.01.001

Pester, M., Rattei, T., Flechl, S., Griebler, C., and Wagner, M. (2010). Impact of soil properties on the composition of bacterial communities in agricultural soils. *Appl. Soil Ecol.* 45, 1–8. doi: 10.1016/j.apsoil.2010.02.003

Prasad, R., and De Vries, F. T. (2019). The role of beneficial fungi in nematode control and soil health. *Front. Microbiol.* 10, 2021. doi: 10.3389/fmicb.2019.02021

Raaijmakers, J. M., Paulitz, T. C., and Steinberg, C. (2009). Microbial ecology of the rhizosphere. *Soil Biol. Biochem.* 41, 1311–1313. doi: 10.1016/j.soilbio.2009.01.019

Ranjan, P., Yadav, S., Singh, R., Kumar, V., and Sharma, A. (2020). 16S rRNA and ITS sequencing reveal bacterial and fungal communities in soil. *Microbial Ecol.* 79, 509–518. doi: 10.1007/s00248-020-01456-7

Rao, V. R., and Rao, R. S. (2016). Role of Planctomycetes in biogeochemical cycles: Current status and future perspectives. *Front. Microbiol.* 7. doi: 10.3389/fmicb.2016.00970

Roe, S., and Owens, M. (2017). Nematode infestation levels in North Alabama soils. *J. Agric. Nematol.* 45, 23–30. doi: 10.21307/jofnem-2017-045

Sambrook, J., and Russell, D. W. (2001). *Molecular Cloning: A Laboratory Manual. 3rd ed* (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press). Vol. 13. doi: 10.1101/0879695765.

Sang, H., Park, J. K., and Kim, K. Y. (2019). The impact of biotic factors on the microbial community structure in soil. *Appl. Soil Ecol.* 134, 115–122. doi: 10.1016/j.apsoil.2018.10.008

Santos, F. A., Silva, M. R., Oliveira, J. P., and Costa, L. M. (2018). Bacterial community structure in response to soil nematode invasion: A focus on rhizosphere interactions. *Appl. Soil Ecol.* 126, 21–30. doi: 10.1016/j.apsoil.2017.11.015

Schliep, K. P. (2011). phangorn: phylogenetic analysis in R. *Bioinformatics* 27, 592–593. doi: 10.1093/bioinformatics/btq706

Shade, A., Caporaso, J. G., Handelsman, J., Knight, R., and Fierer, N. (2012a). Historical links between bacterial diversity and ecosystem function. *Nat. Rev. Microbiol.* 10, 507–517. doi: 10.1038/nrmicro2839

Shade, A., Jacques, M. A., and Barret, M. (2012b). Fundamentals of microbial community assembly and ecology. *Environ. Microbiol.* 14, 4–13. doi: 10.1111/j.1462-2920.2011.02558.x

Shang, Y., and Wang, J. (2022). Plant-parasitic nematode effects on the structure and function of rhizosphere microbiomes. *Environ. Microbiol.* 24, 4732–4746. doi: 10.1111/1462-2920.16001

Siddiqi, M. R. (2000). Nematodes of the genus tylenchus. *J. Nematol.* 32, 1–15. doi: 10.21307/jofnem-2000-032

Siddiqui, Z. A., and Aziz, S. (2024). Plant parasitic nematode-fungus interactions: Recent concepts and mechanisms. *Plant Physiol. Rep.* 29, 37–50. doi: 10.1007/s40502-023-00762-4

Sievert, C. (2020). *Interactive Web-Based Data Visualization with R, plotly, and shiny* (Boca Raton, FL: CRC Press). doi: 10.1201/9780429447273

Singh, R., Kumar, V., Sharma, P., and Gupta, R. (2020). Role of soil microbes in controlling plant-parasitic nematodes. *Appl. Soil Ecol.* 147, 103388. doi: 10.1016/j.apsoil.2019.103388

Smith, J., and Lee, A. (2023). Rhizospheric soil sampling techniques for crop studies. *J. Soil Sci.* 58, 45–58. doi: 10.1016/j.soil.2023.01.001

Tahseen, Q., and Clark, I. M. (2014). Attraction and preference of bacteriophagous and plant-parasitic nematodes towards different types of soil bacteria. *J. Natural History*. 48, 1485-1502. doi: 10.1080/00222933.2013.873088

Thermo Fisher Scientific (2023). NanoDrop™ 1000 Spectrophotometer User Manual (Wilmington, DE: Thermo Fisher Scientific). Available at: https://www.thermofisher.com (Accessed November 4, 2023).

Thomas, S. H., Bowers, L. A., and McKenzie, S. P. (2019). Soil nematode populations in agricultural soils of North Alabama: A regional survey. *South. J. Agric. Sci.* 41, 122–130. doi: 10.21307/sjas-2019-041

Topalović, O., Hussain, M., and Heuer, H. (2020). Plants and associated soil microbiota cooperatively suppress plant-parasitic nematodes. *Front. Microbiol.* 11, 313. doi: 10.3389/fmicb.2020.00313

Van der Heijden, M. G. A., Martin, F. M., Selosse, M. A., and Sanders, I. R. (2015). Mycorrhizal fungi as drivers of ecosystem processes in the rhizosphere. *Science* 349, 1143–1146. doi: 10.1126/science.aab0510

van der Putten, W. H., and Bakker, E. J. (2018). Interactions of plants with nematodes and soil microorganisms. *Trends Plant Sci.* 23, 338–352. doi: 10.1016/j.tplants.2018.01.005

Vasquez, A., Arce, R., and Giordano, G. (2019). Ecological significance of the family Gaiellaceae in soil environments. *Front. Microbiol.* 10. doi: 10.3389/fmicb.2019.01834

Verschoor, B. C. (2002). Carbon and nitrogen budgets of plant-feeding nematodes in grasslands of different productivity. *Appl. Soil Ecol.* 20, 15–25. doi: 10.1016/S0929-1393(02)00010-0

Wagg, C., Bender, S. F., Widmer, D., and van der Heijden, M. G. A. (2014). Soil biodiversity and its benefits for ecosystem functioning: A review. *Nat. Plants* 1, 1–9. doi: 10.1038/nplants.2014.62

Wang, Q., Garrity, G. M., Tiedje, J. M., and Cole, J. R. (2007). Naive Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. *Appl. Environ. Microbiol.* 73, 5261. doi: 10.1128/AEM.00062-07

Wang, Y., Zhang, Y., and Chen, L. (2019). Chloroflexi and its role in soil carbon cycling. *Soil Biol. Biochem.* 131, 48–54. doi: 10.1016/j.soilbio.2019.06.013

Wang, C., Liu, H., Pham, T. C., Hu, X. J., Liu, L., Foba, C. N., et al (2022). Volatiles and hormones mediated root-knot nematode induced wheat defense response to foliar herbivore aphid. *Science of the Total Environment* 815, 152840. doi: 10.1016/j.scitotenv.2021.152840

Wang, Q., Zhang, J., and Zhang, J. (2023). Adaptive responses of soil bacteria to nematode infestation: A meta-analysis. *Environ. Microbiol. Rep.* 15, 15–24. doi: 10.1111/1758-2229.13244

White, T. J., Bruns, T. D., Lee, S. B., and Taylor, J. W. (1990). "Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics," in *PCR Protocols: A Guide to Methods and Applications*, vol. 18, 315–322. San Diego, CA: Academic Press. doi: 10.1016/B978-0-12-372180-8.50042-1

Wickham, H. (2007). Reshaping data with the reshape package. *J. Stat. Software* 21, 1–20. doi: 10.18637/jss.v021.i12

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis* (New York, NY: Springer-Verlag New York) Vol. 1, pp. 1-260. doi: 10.1007/978-3-319-24277-4

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., et al. (2019). Welcome to the tidyverse. *J. Open Source Software* 4, 1686. doi: 10.21105/joss.01686

Wilson, K. (2020). "Preparation of agarose gels for electrophoresis," in *Practical Applications of Electrophoresis: Methods in Molecular Biology* (Springer, New York, NY), 19–26.

Wright, E. S. (2015). DECIPHER: harnessing local sequence context to improve protein multiple sequence alignment. *BMC Bioinf.* 16, 322. doi: 10.1186/s12859-015-0749-z

Wu, J., Wang, D., and Zhang, X. (2019). Effects of soil microbes on plant health and productivity. *Agron. Sustain. Dev.* 39, 9. doi: 10.1007/s13593-019-0583-1

Xie, Y., Wang, Y., and Zhang, X. (2023). Interactions between rhizosphere microbiome and plant-parasitic nematodes: Implications for sustainable pest control strategies. *Soil Biol. Biochem.* 173, 108753. doi: 10.1016/j.soilbio.2023.108753

Yergaliyev, T. M., Alexander-Shani, R., Dimeretz, H., Pivonia, S., Bird, D.M., Rachmilevitch, S., et al. (2020). The bacterial community structure dynamics in Meloidogyne incognita infected roots and its role in worm-microbiome interactions. *Msphere.* 5 (4), 10-1128. doi: 10.1101/2020.03.25.007294

Youssef, N. H., Elshahed, M. S., and McInerney, M. J. (2015). Insights into the evolutionary history of Actinobacteria based on 16S rRNA gene sequences. *Front. Microbiol.* 6. doi: 10.3389/fmicb.2015.00530

Yuan, X., Zhang, X., Liu, X., Han, H., Shen, J., and Zhang, L. (2020a). Impact of root-knot nematode infestation on the diversity and structure of the rhizosphere microbiome in cotton and soybean. *Soil Biol. Biochem.* 147, 107860. doi: 10.1016/j.soilbio.2020.107860

Yuan, J., Liang, Y., and Liu, C. (2020). Impact of plant-parasitic nematodes on soil microbial communities in cotton. *Soil Biol. Biochem.* 148, 107888. doi: 10.1016/j.soilbio.2020.107888

Zhalnina, K., Louie, K. B., Hao, Z., Mansoori, N., da Rocha, U. N., Shi, S., et al. (2022). A reproducible and tunable synthetic soil microbial community provides new insights into microbial ecology. *mSystems* 7, e00951–e00922. doi: 10.1128/mSystems.00951-22

Zhang, X., Liu, Y., Wang, J., Chen, H., Li, Q., and Zhao, L.. (2022). Impact of root-associated fungi on the microbiome structure and diversity in response to nematode stress. *Microbial Ecol.* 83, 482–495. doi: 10.1007/s00248-021-01834-2

Zhang, D., Ma, Y., and Cheng, Y. (2018). The potential roles of Gaiellaceae in soil carbon cycling: Insights from the microbiome. *Soil Biol. Biochem.* 83, 105–111. doi: 10.1016/j.soilbio.2014.12.014

Zhang, W., Wang, M., and Lee, C. Y. (2019). Impact of nematode feeding on microbial diversity and its functional implications in agricultural soils. *Appl. Soil Ecol.* 136, 1–10. doi: 10.1016/j.apsoil.2018.11.008

Zhang, J., Wang, Z., and Liu, Y. (2020). Microbial community structure and function in the rhizosphere: implications for plant growth and soil health. *Soil Biol. Biochem.* 149, 107910. doi: 10.1016/j.soilbio.2020.107910

Zhang, L., Wang, Y., and Zhao, Y. (2023). Microbial community dynamics in response to root-knot nematodes: Implications for sustainable agriculture. *Appl. Microbiol. Biotechnol.* 107, 509–520. doi: 10.1007/s00253-022-12488-6

Zhao, L., Zhang, X., Liu, Y., Wang, J., and Chen, H. (2017). Nematode management in the rhizosphere: Microbial control and interactions. *Front. Microbiol.* 8, 463. doi: 10.3389/fmicb.2017.00463

Zhao, S., Liu, S., and Cheng, W. (2018). Nematode infestation-induced alterations in microbial community structure and their impact on soil health. *Appl. Soil Ecol.* 130, 142–151. doi: 10.1016/j.apsoil.2018.05.010

Zhao, D., Yang, Z., and Xu, L. (2021). Effects of nematode infestation on fungal community dynamics in soil. *Microbial Ecol.* 81, 71–82. doi: 10.1007/s00248-020-01669-8

Zheng, J., Dini-Andreote, F., Luan, L., Geisen, S., Xue, J., Li, H., et al. (2022). Nematode predation and competitive interactions affect microbe-mediated phosphorus dynamics. *Mbio.* 13 (3), e03293-21. doi: 10.1128/mbio.03293-21

Zong, Y., Zhang, C., and Li, J. (2021). Responses of fungal communities to different soil management practices in a coastal saline-alkaline soil. *Appl. Soil Ecol.* 158, 103785. doi: 10.1016/j.apsoil.2020.103785Bais

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 27 March 2025
DOI 10.3389/fpls.2025.1544010

Check for updates

OPEN ACCESS

EDITED BY
Somashekhar M. Punnuri,
Fort Valley State University, United States

REVIEWED BY
Venkatesh Bollina,
Agriculture and Agri-Food Canada (AAFC), Canada
Sumit Jangra,
University of Florida, United States

*CORRESPONDENCE
Wilber Wambi
wambiwilber@gmail.com
Dan Makumbi
D.Makumbi@cgiar.org

RECEIVED 12 December 2024
ACCEPTED 24 February 2025
PUBLISHED 27 March 2025

CITATION
Wambi W, Makumbi D, Asea G, Zeleke H, Bruce AY, Wakgari M, Kwemoi DB and Prasanna BM (2025) Use of multi-trait principal component selection index to identify fall armyworm (*Spodoptera frugiperda*) resistant maize genotypes. *Front. Plant Sci.* 16:1544010. doi: 10.3389/fpls.2025.1544010

COPYRIGHT
© 2025 Wambi, Makumbi, Asea, Zeleke, Bruce, Wakgari, Kwemoi and Prasanna. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Use of multi-trait principal component selection index to identify fall armyworm (*Spodoptera frugiperda*) resistant maize genotypes

Wilber Wambi [1,2*], Dan Makumbi [3*], Godfrey Asea [4], Habtamu Zeleke [1], Anani Y. Bruce [3], Mulatu Wakgari [1], Daniel Bomet Kwemoi [4] and Boddupalli M. Prasanna [3]

[1]African Centre of Excellence for Climate-Smart Agriculture and Biodiversity Conservation, Haramaya University, Dire Dawa, Ethiopia, [2]Bulindi Zonal Agricultural Research and Development Institute, National Agricultural Research Organization, Hoima, Uganda, [3]Global Maize Program, International Maize and Wheat Improvement Center (CIMMYT), Nairobi, Kenya, [4]National Crops Resources Research Institute, National Agricultural Research Organization, Namulonge, Uganda

The Fall armyworm (FAW), *Spodoptera frugiperda* (J. E. Smith) invaded sub-Saharan Africa (SSA) in 2016 and has since become prevalent in many countries, causing significant maize grain yield losses and reduced grain quality. Breeding for host plant resistance to FAW requires improving multiple traits, complicating selection. This study evaluated the use of principal component (PC)-based multi-trait selection indices to identify FAW resistant maize genotypes. A total of 192 maize hybrids alongside four commercial hybrids, were evaluated over four seasons under artificial FAW infestation. Data on FAW leaf feeding damage (LD) at 7, 14, and 21 days after infestation, and ear damage (ED), ear rot (ER), and grain yield (GY) were recorded. The data were subjected to analysis of variance and PC analysis, and results used to construct two economic weight-free selection indices: PC1-based index (PC1BI) and PC2-based index (PC2BI). Broad-sense heritability estimates were 0.59 to 0.73 for LD, and 0.69 for GY. The two PCs explained 97.1% of the variation among the hybrids. PC1BI, with higher loadings for the leaf feeding damage traits, showed the larger desired gains for these traits (–2.92 to –3.84%) and GY (19.9%), making it a superior index to PC2BI. PC1BI identified six promising hybrids with GY above the cutoff of 7.0 t ha^{-1} for selection under FAW infestation. PC2BI exhibited larger gains for ED (–11.1%) and ER (–45.4%). The index-based selected hybrids consistently outperformed the commercial hybrid checks. The PC-based indices have the potential to serve as valuable tools for breeders to maximize selection gains; however, modifications are necessary to incorporate other desirable agronomic and adaptive traits.

KEYWORDS

Fall armyworm, host plant resistance, multi-trait selection, principal component analysis, selection gain, *Spodoptera frugiperda*, weight-free selection index

Introduction

Maize is an important and strategic crop in sub-Saharan African (SSA) for food, feed, and plays key roles in the national economies of several countries as a key industrial and export commodity (Badu-Apraku and Fakorede, 2017; Ekpa et al., 2018, 2019; Erenstein et al., 2022; UBOS, 2022). Sustained production of maize is curtailed by several factors including declining soil fertility, climate change, and emerging pests and diseases (Sisay et al., 2019; Prasanna et al., 2021; Asfaw et al., 2024). The outbreak of the invasive fall armyworm (FAW) pest first reported in 2016 is among the most important pest limiting the production of maize in SSA (Goergen et al., 2016; Otim et al., 2018; Sisay et al., 2019; Tambo et al., 2020). The pest has a broad host range, but its impact is severe on maize fields causing plant damage and yield losses ranging between 20 to 70% especially during aggressive feeding larval stage (Day et al., 2017; Ramirez-Cabral et al., 2017; Montezano et al., 2018). The larval feeding causes several damages to the leaves, stems, and maize ears, further exposing cobs to ear rots and mycotoxin contamination, reducing the quality and safety of grains and their products (FAO, 2017; Pruter et al., 2020; Palumbo et al., 2020; Logrieco et al., 2021). The impact of the pest is mainly severe for smallholder farmers with limited access to inputs, especially pesticides that are costly and require training on use to limit exposure and pesticide risks on humans, crops, livestock and agro-biodiversity (Day et al., 2017; Ramirez-Cabral et al., 2017; Kassie et al., 2020; Tambo et al., 2020; Yan et al., 2022).

The FAW is difficult to control and manage due to its migratory, transboundary and polyphagous behavior, short life cycle and nature of mostly small holder farming in SSA (FAO, 2021; Otim et al., 2021). Host resistance provides a cost-effective, user-friendly, and environmentally safe component of an integrated pest management strategy against FAW under smallholder conditions (Prasanna et al., 2022). Most tropical breeding programs are still in their early stages of host resistance development, and only a few first-generation FAW-tolerant hybrids are commercially available, thus limiting the use of host resistance (Prasanna et al., 2022). Breeding for next-generation and durable resistance demands the consideration of several traits in selection and advancement processes. The key traits targeted by breeders include foliar damage and grain yield traits that benefit the producer, as well as those that influence product quality and consumer acceptance, such as ear damage and ear rot. However, FAW-inflicted leaf-feeding damage and ear quality traits often show an unfavorable correlation with grain yield (Kasoma et al., 2021; Kamweru et al., 2022; Prasanna et al., 2022). Selecting uncorrelated traits focusing on direct selection approaches makes genetic improvement challenging for multiple traits (Olivoto and Nardino, 2020; Zuffo et al., 2020; Batista et al., 2021; Ambrósio et al., 2024). Investigating selection indices aimed at selecting desired second-generation host resistant varieties is crucial given the uncertainty when selection is focused on direct trait approaches. Studies have demonstrated that employing selection indices for multiple traits offers successful chances for breeding several traits into a single genotype (Lopez-Cruz and de los Campos, 2021; Padjung et al., 2021; Anshori et al., 2022; Cerón-Rojas and Crossa, 2022; Farid et al., 2022; Anshori et al., 2024). Breeders utilize a variety of selection indices, such as the Smith-Hazel index (Smith, 1936), base index (Williams, 1962), restricted selection index (Kempthorne and Nordskog, 1959), and the Tai index (Tai, 1977). Some of these indices require the use of economic weights but economic weights for traits are subjective and not easy to determine. Selection indices that require weights generated intrinsically from data are easy to use and amenable for breeders. An example of such an index is the principal component (PC) based index which is possibly the most used (Agudelo-Gómez et al., 2016; Tang et al., 2021; Padjung et al., 2021). Principal component analysis is useful in eliminating redundancy in univariate analyses (Iezzoni and Pritts, 1991), and therefore, PC-based indices also eliminate multicollinearity among variables. The PC-based index method has been used to identify maize varieties with acceptable nutritional value and promising physiological seed quality traits (Tang et al., 2021; Padjung et al., 2021). Principal component indices could thus be useful when selecting FAW-resistant genotypes since some of the traits are scored multiple times, and others show unfavorable correlations. The objective of this study was to evaluate PC-based indices for the selection of maize hybrids with resistance to FAW and improved grain yield under artificial infestation conditions.

Materials and methods

Plant materials

In 2019, nineteen (19) inbred lines were crossed in a diallel mating scheme to generate 171 F1 test hybrids without reciprocals. The inbred lines were developed by CIMMYT and showed promising host responses to FAW feeding damage under artificial infestation in preliminary tests that were done in 2017 and 2018. Table 1 details characteristics, pedigree, and origin of parental lines utilized in the half-diallel crosses. The crosses were made at Kenya Agricultural and Livestock Research Organization (KALRO) Kiboko Research Station (2°15'S, 37°75'E, 975 m a.s.l) in Kenya. In addition to the diallel hybrids, 17 experimental hybrids, including two internal genetic gain checks (CKH10769 and CKH10717), and two commercial check hybrids (WH401 and WH505) were included in the study.

Evaluation of response to FAW infestation

A total of 171 diallel hybrids were evaluated along with 15 or 17 experimental hybrids, two internal genetic gain checks (CKH10769 and CKH10717), and two commercial check hybrids (WH401 and WH505) for four seasons under artificial FAW infestation. In 2020, the trial included 192 hybrids, while in 2021, there were 190 hybrids. The experimental design was a 4 × 48 alpha-lattice in 2020 and a 5 × 38 alpha-lattice in 2021, each with two replications. The experiments were conducted in screen houses in Kiboko, Kenya. The hybrids were planted in a single-row plot, 3 meters long, with 0.75 meters between rows and 0.25 m between plants. FAW neonates, obtained from the CIMMYT insectary at KALRO Katumani Agricultural Experimental Station in Machakos County,

TABLE 1 List and characteristics of parental lines utilized in the half diallel crossing.

Code	Name	Pedigree	Characteristics	Origin
1	CML247	(G24-F119/G24-F54)-6-4-1-1-B-B-B-B	Tolerant to ear rot	Mexico
2	CML269	P25STE-C1-FS13-6-1-1-#-B-B-B-B	Tolerant to ear rot	Mexico
3	CML17	P22TSR-B-B-40-2-1-2-B-B-B-B	Lodging resistant	Mexico
4	CML23	P25-FS112-2-2-2-1-4-B-B-B-B	Tolerant to ear rot	Mexico
5	CML274	(AC7643/P43-F7)-2-3-4-3-B-B-B-B	Tolerant to ear rot	Mexico
6	CML371	MBRETW-F2-56-1-1-1-B-B-6-B-B-B-B	Multiple borer resistance	Mexico
7	CML372	MBRETW-F2-177-3-1-1-B-B-B-B	Multiple borer resistance	Mexico
8	CML402	AC8222-6-2-2-B-#-#-2-B*3-1-B-B-B-B	Tolerant to ear rot	Mexico
9	CML476	P21MRRS-C1-525-1-B-B-B-B	Susceptible to FAW	Mexico
10	DL187008	MDRC3 Bc/MBRC5 Bc F59-1-B-#-1-2-B-B-B-B	Multiple borer resistance	Kenya
11	DL187009	MBR/MDRC4 Bc F34-1-B-#-1-1-B-B-B-B	Multiple borer resistance	Kenya
12	DL187010	MIRTC5 Bco F62-2-2-1-1-2-1-B-B-B-B	Multiple insect resistance	Kenya
13	DL187011	P84c3BcxMIRTC5 Bco F10-1-2-2-2-3-1-B-B-B-B	Multiple insect resistance	Kenya
14	DL187012	P84c3BcxMIRTC5 Bco F80-4-2-1-4-1-1-B-B-B-B	Multiple insect resistance	Kenya
15	DL187019	MBRC6 Bc F234-1-B-#-1-1-B-B-B-B	Multiple borer resistance	Kenya
16	CML334	P590-C3-F374-2-1-2-B-#-3-3-B-B-B-B	Southwestern corn borer resistance	Mexico
17	CKSBL10153	[(MIRTC4Am F128-B-1-3-B-B x CL-02450)-B]-1-B-B-B-B	Stem borer tolerance, FAW tolerance	Kenya
18	CML71	ANTGP2-5-#-1-2-1-1-5-5-7-B-B-B-B	Tolerant to FAW	Mexico
19	CML345	P390SCB-C1-F72-1-1-1-1-#-6-B-B-B-B	Tolerant to FAW	Mexico

Kenya, were used to artificially infest the hybrids. The rearing of FAW neonates followed the protocols described in detail by Prasanna et al. (2018) and Anyanda et al. (2022). Seven neonates were placed in the whorl of each plant at the V5 stage using a camel brush, following the protocols outlined by Prasanna et al. (2018).

Data collection

Fall armyworm leaf feed damage (LD) was assessed at 7 (LD1), 14 (LD2), and 21 (LD3) days after artificial infestation using a visual rating scale of 1–9 (Davis and Williams, 1992; Prasanna et al., 2018). A score of 1 is considered highly resistant, 2 to 3 resistant, 4 to 5 moderately resistant, 6 to 7 susceptible, and 8 to 9 highly susceptible. The average leaf feeding damage score (LD_AV) was computed from the three ratings. At harvest, ear damage (ED) was also rated on a scale of 1–9 (Prasanna et al., 2018; Kamweru et al., 2022), where 1 indicated no ear feeding damage and 9 represented severe ear feeding damage. Ear rot (ER) was recorded by counting rotten ears per plot after harvest, expressed as a percentage of the total harvested ears. All the harvested ears were then weighed to determine plot field weight data. Grain moisture content was measured using a moisture meter. Using field weight, grain moisture content, and an average shelling percentage of 80%, grain yield (GY, t ha^{-1}) was calculated and adjusted to 12.5% moisture content.

Data analyses

Analysis of variance

Data were checked for normality using the Shapiro-Wilk test by examining residual plots, and histograms. The data were found to be normally distributed, and therefore no data transformation was required. A linear mixed model was used to perform analyses of variance (ANOVA) across environments using the META-R software (Alvarado et al., 2020). The linear mixed model used for the analysis of the data was:

$$Y_{ijkl} = \mu + G_i + E_j + R_k(E_j) + B_l(ER)_{jk} + GE_{ij} + \varepsilon_{ijkl}$$

where Y_{ijk} is the response variable; μ is the intercept; G_i is the effect of the ith genotype; E_j is the effect of the jth environment; $R_k(E_j)$ is the effect of kth replicate in the jth environment; $B_l\ (ER)_{jk}$ is the effect of the lth block within the kth replicate at the ith environment; $GE_{ij}$is the effect of the interaction between the ith genotype and the jth environment; and ε_{ijkl} is the experimental error associated with the i*th* genotype, j*th* environment, k*th* replicate and l*th* block. The error is assumed to have a mean zero and a homoscedastic variance σ^2_ε. The best linear unbiased estimates (BLUEs) computed from a combined ANOVA were used to create the indices and perform correlation analysis. Broad-sense heritability was estimated for combined environments using variance components following Hallauer et al. (2010) as:

$$H^2 = \frac{\sigma_G^2}{\sigma_G^2 + \frac{\sigma_{GE}^2}{e} + \frac{\sigma_E^2}{er}}$$

where σ_G^2 is the genotypic variance, σ_{GE}^2 is the genotype × location interaction variance, e is the number of environments, r is the number of replicates, and the σ_E^2 is the residual variance.

Construction of principal component indices for hybrid selection

The data for six FAW-resistance-related parameters (LD1, LD2, LD3, ED, ER, and GY) from the 192 (2020) or 190 (2021) maize genotypes were standardized to have a mean of zero and variance of one for each trait. The standardized values were then subjected to Principal Component (PC) analysis to obtain weights for constructing the selection indices using the META-R software. The decision on the number of PCs to use in the index was informed by analysis of the scree plot, where only PCs with eigenvalues greater than one were selected (Supplementary Figure 1). This criterion led to selection of the first two principal components (PC1 and PC2) used in the study. The index score values for each genotype were calculated by summation of the products of the BLUEs across environments for each trait and their respective weights (eigen vectors) on the first two principal component axes (PC1 and PC2) following the procedures outlined by Tang et al. (2021). The computations were done in Microsoft Excel, considering a set of all six FAW resistance-related traits. Two PC-based selection index schemes were developed: a PC1-based index (PC1BI) and a PC2-based index (PC2BI).

The model for constructing PC1BI was:

$$\text{PC1BI} = -0.95 \times \text{LD1} + -0.93 \times \text{LD1} + -0.91 \times \text{LD3} + -0.50 \times \text{ED} + -0.70 \times ER + 1.00 \times \text{GY}$$

while the model used to construct PC2BI was:

$$\text{PC2BI} = -0.36 \times \text{LD1} + -0.44 \times \text{LD2} + -0.45 \times \text{LD3} + 1.00 \times \text{ED} + 0.83 \times \text{ER} + -0.08 \times GY$$

The LD1, LD2, LD3, ED, ER, and GY correspond to the hybrid means for leaf feeding damage score at 7, 14, and 21 days after infestation, ear damage score, ear rot score, and grain yield, respectively. The multiplicative factors for each trait were the eigen vectors from the PC analysis. The genotypic correlations among traits and the index score values were computed and visualized using the R package '*psych*' version 2.4.6 (William, 2024).

Performance of PC-indices and hybrid selection

To evaluate the selection efficiency of the PC-based indices, multi-trait selection gains among seven FAW resistance-related traits either through conventional single-trait or principal component-based index selection options were compared. Each trait, including LD1, LD2, LD3, LD_AV, ED, ER, and GY, was considered as an independent conventional single-trait direct selection option. In each selection scheme, the hybrids were ranked and the average of the 15 or 10 top-performing selected hybrids was considered while estimating the selection deferential for the traits. The selection gain (GS) for each trait was calculated as a percentage of the selection differential, expressed relative to the grand mean of the 192 hybrids, and then multiplied by the trait heritability (Santos et al., 2018; Zuffo et al., 2020; Norman et al., 2022). Accordingly, the equation used to estimate selection gain (GS) was:

$$GS = \frac{(Xs - Xo) \times H^2}{Xo}$$

where, Xs is mean of the 15 or 10 top-performing selected hybrids, Xo is the grand mean of the 192 hybrids, $(Xs - Xo)$ = selection deferential, and H^2 is the broad-sense heritability of the trait.

Additionally, the relative gain, estimated as RG = [(Genotype–Check)/Check] × 100 was used to compare the performance of the 15-best index-selected hybrids against the commercial checks WH401 and WH505 following the method of Matias et al. (2020). A grain yield of 7 t ha^{-1} was used as the cutoff, based on the trait metrics developed and used in the Stage gate advancement process by the Global Maize Program of CIMMYT for Product Profile #1 (intermediate maturity maize). Relationship heatmaps based on the top 15 and five bottom index-selected hybrids were generated using the 'pheatmap' package in R to visualize and identify superior genotypes that exhibited desired combinations for several traits (Kolde, 2019).

Results

Analysis of variance, variance components, heritability and genotype mean performance

A combined ANOVA revealed significant ($P < 0.01$) environment (E), genotype (G), and G × E mean squares for all traits (Table 2). The genotypic variance (σ_G^2) and G × E variance (σ_{GE}^2) were of similar magnitude for LD1 but σ_G^2 was larger than σ_{GE}^2 for the other traits (Table 3). The broad-sense heritability (H^2) estimates were moderate for LD1, LD3, ED, ER, and GY (range 0.59 to 0.69) and high (range 0.73 and 0.79) for LD2 and LD_AV. The mean LD across the three scores was 5.1 while the mean GY was 5.1 t ha^{-1} for the trial (Table 3). The performance of the test hybrids and checks is presented in Supplementary Table 1. A total of 186 hybrids had leaf feeding damage 1 (LD1) scores of<4, while six hybrids showed LD1 scores >4. For LD2, moderate resistance scores ranged from 5.4 (E163) to 5.9 (E82) for 66 hybrids, while for LD3 the scores ranged from 5 (E152) to 5.9 (E94) for 148 hybrids. Fifty-four experimental hybrids recorded average leaf feeding damage

TABLE 2 Combined analysis of variance for FAW resistance parameters and agronomic traits of 192 hybrids evaluated under artificial FAW infestation in 2020-2022.

Source of variation	df	LD1	LD2	LD3	LD_AV	ED	ER	GY
Environment (E)	3	42.82***	38.08***	168.89***	20.16***	88.33***	4023.00***	982.60***
Rep (E)	4	9.47***	18.97***	18.33***	6.70***	4.09***	420.00***	60.90***
Block (Rep/E)	336	0.24***	0.45***	0.68***	0.25***	0.30***	33.00**	4.50***
Genotypes	191	0.46***	0.59***	0.88***	0.42***	0.62***	86.00***	11.60***
Genotypes × E	565	0.22***	0.28***	0.47***	0.16***	0.30***	40.00***	4.30***
Residual	422	0.12	0.11	0.27	0.70	0.17	25.00	1.70

, * Significant at $P < 0.01$ and $P < 0.001$, respectively.
LD1, LD2, and LD3, leaf feeding damage score at 7, 14, and 21 days after infestation; LD_AV, Average leaf feeding damage score; ED, Ear damage; ER, Ear rot; GY, Grain yield.

(LD_AV)<5, with check hybrids WH401 and WH505 showing LD_AV >6. For ear damage, all the genotypes recorded ear damage scores<3. Of the test genotypes, 141 experimental hybrids had ER<10%, while the check entries exhibited ER between E192 (17.4%) to E191 (31.3%). Grain yield for the experimental hybrids under artificial FAW infestation ranged from 2.2 to 7.8 t ha^{-1}, with 11 hybrids yielding ≥ 7 t ha^{-1}. The best check WH505 yielded 4.4 t ha^{-1}.

Principal component analysis, correlations among FAW resistance parameters, and index scores

Principal component analysis revealed that the first two PCs explained 97.1% of the total variation among the hybrids (Table 4; Figure 1). The traits with the highest loadings on PC1 (64.2%) were the leaf-feeding damage scores and grain yield. The traits with the highest loadings on PC2 (32.9%) were ear damage and ear rot. The genotypic correlations between the time series FAW leaf feeding damage traits LD1, LD2, LD3, and LD_AV were strong (r = 0.3 to 0.9) (Figure 2). The PC1BI and PC2BI index values for the hybrids are presented in Supplementary Table 1. The desirable index value is the larger value. The PC1BI index values ranged from -26.3 to -8.7, for WH505 and E154, respectively. For PC2BI, the index values ranged from -27.6 to 1.4 for hybrids WH505 and E28, respectively. The PC1BI and PC2BI index values were strongly correlated with all traits except GY with PC2BI values.

Performance of PC-based indices and hybrid selection

The expected selection gains for FAW resistance parameters and agronomic traits based on the PC indices are presented in Table 5 and Figure 3. Using the PC1BI resulted in the desired genetic gains for all traits for the top 10 and 15 selected hybrids. A similar result was obtained using PC2BI except for GY when the top 15 hybrids were selected. There were similar gains in the desired direction for the traits with high loadings on PC1 with selection using PC1BI. The selection gain for GY was higher (19.9%) under PC1BI compared with PC2BI (1.68%) when selecting the top 10 hybrids. Index PC2BI, based on high PC2 loadings on ED and ER, showed larger gains for these traits relative to PC1BI. Direct selection based on average leaf feeding damage (LD_AV) showed larger gains for LD1, LD2, LD3, and ER compared to PC1BI, direct selection based on ER (ER_DS), and direct selection based on GY (GY_DS). Direct selection based on GY resulted in larger gains for GY but undesirable direction for ED and ER. The ranking of the top 15 and bottom five experimental hybrids under different selection schemes is shown in Table 6. The LD of the top 10 hybrids ranged

TABLE 3 Variance component and broad-sense heritability (H^2) estimates and means for FAW resistance parameters and agronomic traits of 192 hybrids evaluated under artificial FAW infestation in 2020-2022.

Traits	σ^2_G	σ^2_{GE}	σ^2_e	H^2	Mean	Range	$LSD_{0.05}$	CV (%)
LD1	0.03***	0.02***	0.12	0.62	3.52	2.90 – 4.80	0.33	9.71
LD2	0.04 ***	0.01ns	0.11	0.73	6.04	5.40 - 7.30	0.33	5.53
LD3	0.05***	0.00ns	0.27	0.59	5.72	5.00 – 6.90	0.42	9.12
LD_AV	0.04***	0.00ns	0.07	0.79	5.09	4.60 – 6.20	0.27	5.28
ED	0.04***	0.02*	0.18	0.59	1.98	1.40 – 2.90	0.37	21.27
ER	11.03***	3.33	28.80	0.65	8.14	1.10 – 31.30	5.71	65.91
GY	0.88***	0.72***	1.70	0.69	5.11	2.20 – 7.80	1.55	25.55

*, *** Significant at $P < 0.05$ and $P < 0.001$, respectively. σ^2_G, genotypic variance; σ^2_{GE}, variance of genotype x environment interactions; σ^2_e, error variance; LD1, LD2, and LD3, Leaf feeding damage score at 7, 14, and 21, days after infestation; LD_AV, Average leaf feeding damage score; ED, Ear damage; ER, Ear rot; GY, Grain yield.

10.3389/fpls.2025.1544010

TABLE 4 Eigenvectors of the first two principal component axes (PC1 and PC2) based on correlation matrix of FAW resistance parameters.

Traits	PC1	PC2
Leaf feeding damage score at 7 days (LD1)	**-0.97**	0.10
Leaf feeding damage score at 14 days (LD2)	**-0.98**	0.25
Leaf feeding damage score at 21 days (LD3)	**-0.93**	0.30
Ear damage	-0.19	**-1.00**
Ear rot	-0.39	**-0.94**
Grain yield	**1.00**	0.05
Proportion of variance explained	64.2%	32.9%

Bold values indicate loadings above 50% on each PC.

from 4.6 to 5.2, while their grain yield ranged from 5.1 to 7.7 t ha^{-1}. Six of the top 15 based on PC1BI yielded > 7 t ha^{-1}. Out of the top 15 hybrids based on PC1BI, five hybrids (E116, E154, E161, E163, and E74) were ranked among the top 15 hybrids selected based on PC2BI while four hybrids (E141, E148, E163, and E183) were among the top under direct selection based on LD_AV. Hybrid E163 was consistently among the top 15 through all three selection schemes. Two genotypes E155 and E163 were selected under both PC2BI and LD_AV selection schemes. The percentage relative selection gain of each of the 15 top hybrids against the commercial checks WH401 and WH505 under different selection schemes showed that the hybrids would result in larger relative gains compared to the two commercial checks (Table 7). The gains were mainly observed for LD_AV, ED, ER, and GY, with the larger magnitude gains recorded for the PC-based selection indices. The

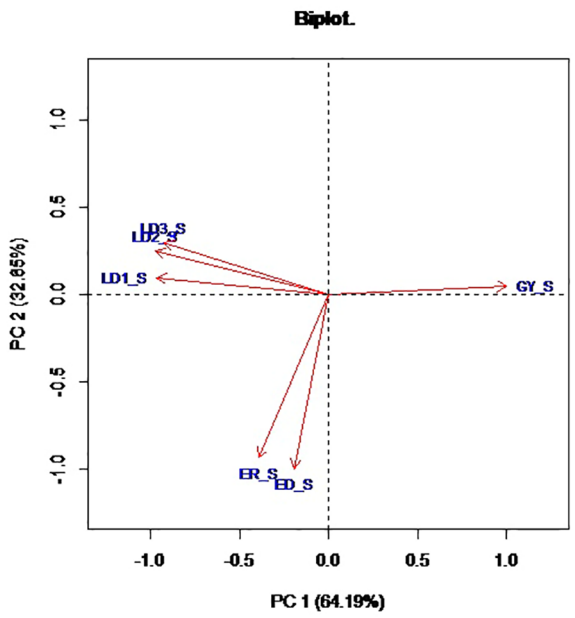


FIGURE 1
Biplot of the first two principal components for FAW resistance parameters and agronomic traits of 192 hybrids. LD1, LD2, LD3 = Leaf feeding damage score at 7, 14, 21 days after infestation, respectively; ED, Ear damage; ER, Ear rot; GY, Grain yield.

relationship heatmaps (Figure 4) categorized the genotypes into two main groups. One cluster included hybrids that clustered together with the check varieties characterized with poor trait-genotype associations, while the other had 15 index-based selected genotypes that were divided into subgroups for each index scheme. Under PC1BI, one subgroup comprised six top yielding hybrids (E78, E159, E115, E133, E141, and E181) that exhibited strong association with GY (Figure 4A). Seven hybrids (E74, E111, E34, E154, E116, E116, and E163) in second subgroup combined relatively higher GY and lower ER and ED damage under the same index. Out of the seven, four hybrids (E116, E154, E161, and E163) combined higher GY and lower LD_AV scores, and one E174 combined higher GY and lower ED scores under PC2BI (Figure 4B).

Discussion

Until 2016, when FAW was reported in Africa (Goergen et al., 2016), it had primarily been a maize pest in the Americas. Its emergence on the African continent prompted the efforts to identify suitable sources of genetic resistance within the available germplasm to develop resistant maize hybrids. These initial breeding efforts, conducted under artificial FAW infestation, led to the identification of several inbred lines with tolerance to FAW (Prasanna et al., 2022). To develop suitable hybrids, these inbred lines must be evaluated in hybrid combination under different management conditions. This process requires consideration of numerous traits to identify the most suitable genotypes. Given the large number of traits to be considered in a breeding strategy, selection of the most relevant ones can be challenging. This process thus requires an efficient approach for selecting hybrids that combine resistance to FAW, good agronomic performance, and a number of essential adaptive traits. Therefore, this study explored the use of principal component-based selection indices to efficiently identify genotypes that combine the desired traits under artificial FAW infestation. The key traits important for FAW resistance breeding in the study included leaf feeding damage, ear damage, ear rot incidence, and grain yield.

Genetic variation and heritability estimates

Significant genotypic variation was observed for all FAW resistance parameters and agronomic traits, indicating substantial genetic variability in this germplasm for breeding for host resistance to FAW. Genetic variability is critical to making genetic progress in breeding. The findings are in agreement with previous investigations on FAW that reported genetic variations for FAW resistance related traits in tropical and temperate maize (Kasoma et al., 2021; Kamweru et al., 2022; Soujanya et al., 2022; Warburton et al., 2023). The development of the insect-resistant populations that are progenitors of some of the lines used in this study was based on a diverse set of Caribbean maize and Tuxpeño accessions from Mexico (Mihm, 1997). This underscores the importance of the large genetic variability in addressing breeding for a pest like FAW in a new environment.

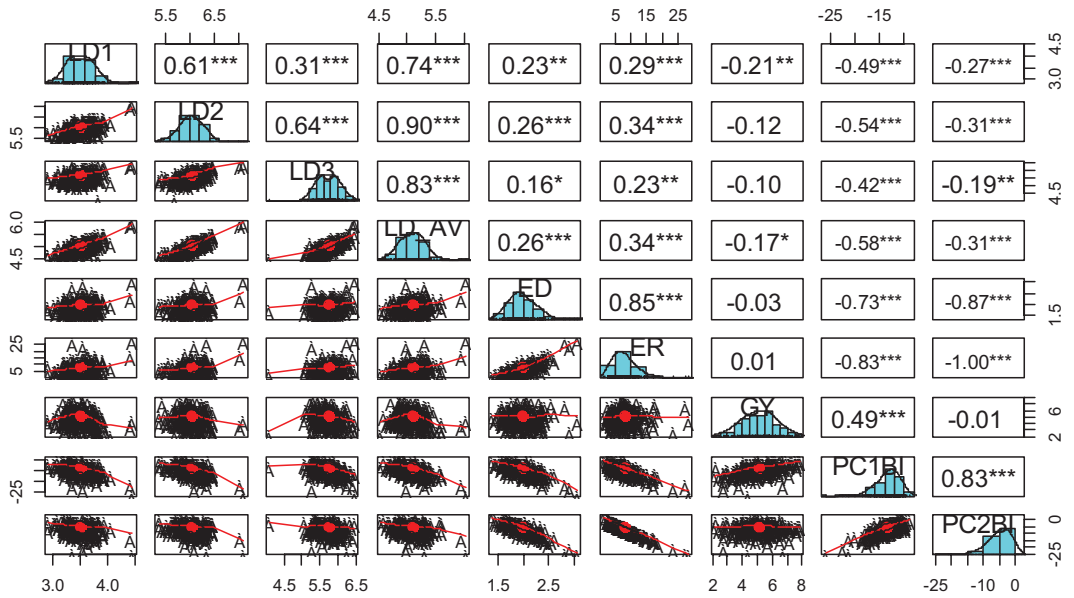


FIGURE 2
Genotypic correlations between FAW resistance parameters and index score values. LD1, LD2, LD3 = Leaf feeding damage score at 7, 14, 21 days after infestation, respectively; ED, Ear damage; ER, Ear rot; GY, Grain yield; PC1BI, Principal component 1-based index; PC2BI, Principal component 2-based index. *, **, *** Significant at $P < 0.05$, $P < 0.01$, and $P < 0.001$, respectively.

In the present study, the environment and G × E interaction were significant, suggesting that the different seasons of evaluation could have led to differential genotype responses. Similar results were reported in earlier studies on FAW (Widstrom et al., 1992; Kasoma et al., 2021; Kamweru et al., 2022; 2023). Differences in temperature over testing seasons are known to influence the biological activities of insects including FAW (Ramirez-Cabral et al., 2017; Schneider et al., 2022; Skendžić et al., 2021; Yan et al., 2022). In plant breeding programs, traits with high heritability are more likely to be improved rapidly (Nyquist, 1991). This study revealed that broad-sense heritability estimates for the FAW resistance parameters were moderate to high, which suggests that breeding for FAW resistance using this germplasm could lead to reasonable genetic gains from selection for these traits. The high broad-sense heritability estimates also suggest that narrow-sense heritability for these FAW-related parameters would be moderate to high (Falconer and Mackay, 1996). Previous studies on FAW have also reported moderate to high broad-sense heritability estimates for FAW resistance parameters in maize (Kasoma et al., 2021; Kamweru et al., 2022; 2023).

Performance of PC-based indices and hybrid selection

A selection index offers a criterion for selection that objectively assesses the genotypic values of individuals or families (Subandi et al., 1973). An effective selection index requires strong genetic correlations among the traits that are included in an index (Oloyede-Kamiyo, 2019; Olivoto and Nardino, 2021; Batista et al., 2021). In this study, the genotypic correlations among the FAW resistance traits were strong, justifying their inclusion in the selection index. Similarly, strong genotypic correlations between the leaf-feeding damage traits, ED, and ER were reported by Kamweru et al. (2022). One of the challenges of using selection indices is the complexity of assigning economic weights (e.g. Smith, 1936; Smith et al., 1981). Several authors have suggested several economic weight-free indices (Williams, 1962; Subandi et al., 1973; Stromberg and Compton, 1989; Tang et al., 2021). Two economic weight-free PC-based selection indices were developed in this study, which together explained 97% of the variation in this germplasm. The first index, PC1BI explained largely the variation in the three leaf feeding damage traits and GY, whereas PC2BI explained the variation in ear-related traits ED and ER. When both indices were used to select the top 15 hybrids, the checks were not among the top hybrids selected. This was because the commercial check hybrids WH401 and WH505 performed poorly in relation to the key traits for FAW resistance with high loadings on both indices, but it also points to the good discriminatory power of the two indices developed. Tang et al. (2021) demonstrated that using multivariate selection tools, such as PCA, is an effective method for combining multiple traits into a single maize genotype.

The use of PC1BI to select the top 10 or 15 hybrids resulted in balanced gains in the desired direction for all traits, with the largest genetic gain for GY of 19.9% (top 10) and 18.8% for top 15%. Interestingly, the improvements in GY achieved using PC1BI were nearly identical to those obtained by directly selecting for GY. Oloyede-Kamiyo (2019) also reported favorable selection gains for GY in a study evaluating the efficiency of four index-based selection methods for resistance to the lepidopteran stem borer *Chilo partellus*. Several other studies have highlighted the beneficial application of selection indices to identify high-yielding hybrids

10.3389/fpls.2025.1544010

TABLE 5 Principal component-based index (PC1BI and PC2BI) and direct selection gains and direction of selection scenarios for FAW resistance parameters and agronomic traits.

Selection scheme	Traits	Selection gains	Direction	Selection gains	Direction
		Top 10 selected hybrids		Top15 selected hybrids	
PC1BI					
	LD1	-3.43	+	-2.95	+
	LD2	-2.92	+	-2.14	+
	LD3	-3.04	+	-2.66	+
	LD_AV	-3.84	+	-3.13	+
	ED	-5.68	+	-6.62	+
	ER	-27.77	+	-27.89	+
	GY	19.90	+	18.83	+
PC2BI					
	LD1	-2.81	+	-1.69	+
	LD2	-1.84	+	-0.96	+
	LD3	-1.66	+	-0.94	+
	LD_AV	-2.43	+	-1.33	+
	ED	-11.05	+	-10.08	+
	ER	-45.43	+	-42.43	+
	GY	1.68	+	-0.63	–
LD_AV_DS					
	LD1	-5.33	+	-5.51	+
	LD2	-5.08	+	-4.63	+
	LD3	-5.21	+	-4.49	+
	LD_AV	-6.46	+	-5.92	+
	ED	-4.31	+	-2.86	+
	ER	-21.58	+	-15.86	+
	GY	-4.64	–	-4.92	–
ED_DS					
	LD1	-0.45	+	-1.15	+
	LD2	-0.09	+	-0.76	+
	LD3	-0.99	+	-1.06	+
	LD_AV	-0.71	+	-1.21	+
	ED	-13.53	+	-12.77	+
	ER	-36.84	+	-33.84	+
	GY	2.60	+	1.62	+
ER_DS					
	LD1	-2.81	+	-1.85	+
	LD2	-1.84	+	-1.47	+
	LD3	-1.66	+	-2.08	+
	LD_AV	-2.43	+	-2.19	+

(Continued)

10.3389/fpls.2025.1544010

TABLE 5 Continued

Selection scheme	Traits	Selection gains	Direction	Selection gains	Direction
		Top 10 selected hybrids		Top15 selected hybrids	
ER_DS					
	ED	-11.05	+	-9.05	+
	ER	-45.43	+	-42.85	+
	GY	1.68	+	-2.50	–
GY_DS					
	LD1	-1.51	+	-1.11	+
	LD2	-1.18	+	-0.72	+
	LD3	-1.82	+	-1.84	+
	LD_AV	-1.85	+	-1.54	+
	ED	1.78	–	1.33	–
	ER	0.78	–	3.67	–
	GY	31.65	+	29.17	+

+, Desired; -, Undesirable; PC1BI, Principal component 1-based index; PC2BI, Principal component 2-based index; and LD_AV_DS, ED_DS, ER_DS, and GY_DS = direct selection schemes based on average leaf feeding damage, ear damage, ear rot and grain yield, respectively.
LD1, LD2, LD3, Leaf feeding damage score at 7, 14, 21 days after infestation, respectively; LD_AV, Average leaf feeding damage; ED, Ear damage; ER, Ear rot; GY, Grain yield.

with favorable trait combinations in maize (Widstrom et al., 1982; Mhike et al., 2012; Makumbi et al., 2018; Crispim-Filho et al., 2020). In contrast, a study on sorghum using the functional plant loss index (FPLI) which is akin to PC1BI, did not show a significant relationship with grain yield performance (Singh et al., 2011). With the application of PC2BI, selection gains in the desired direction were achieved for all traits when selecting the top 10 hybrids, although larger gains were observed for ED and ER. While both indices resulted in desirable gains, PC1BI proved to be a better index, with a significantly higher gain for GY (19.90) compared to 1.68 for PC2BI. Direct selection methods for leaf feeding damage, ear damage or GY, resulted in less balanced selection gains for all traits compared to the PC-based indices. This result further underscores the importance of using selection indices in breeding programs focused on maximizing high yield potential and improving tolerance to multiple stresses.

With the application of the PC1BI index, several promising hybrids with grain yield ≥ 7.0 t ha^{-1} and good overall index scores were identified. These hybrids align with the current priority trait metrics for FAW resistance breeding at CIMMYT (Prasanna et al., 2022). This suggests that PC1BI could serve as an excellent selection index tool in FAW resistance breeding to identify genotypes that combined FAW resistance with high yield potential. Similar recommendations for using selection indices have been made for corn ear worm (Widstrom et al., 1982) and *Chilo partellus* resistance in maize (Oloyede-Kamiyo (2019) and sorghum (Singh et al., 2011). Some hybrids selected using the PC2BI index showed reduced ear and rot damage though their yield was slightly below

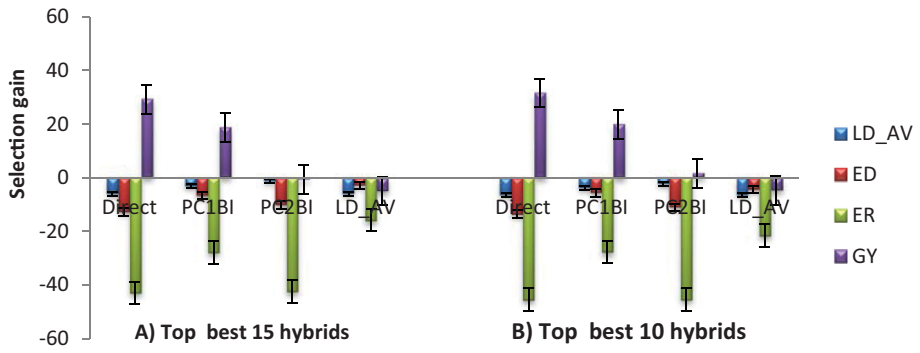


FIGURE 3
Selection gain under PC and direct-based selection schemes based on: **(A)** top 15 **(B)** top 10 selected hybrids. LD_AV, Average leaf feeding damage score; ED, Ear damage; ER, Ear rot; GY, Grain yield.

TABLE 6 Ranking of the top 15 and bottom five experimental hybrids under PC1BI, PC2BI, and LD_AV selection schemes.

Hybrids	LD_AV	ED	ER	GY	Value	Rank	Hybrids	LD_AV	ED	ER	GY	Value	Rank	Hybrids	LD_AV	ED	ER	GY	Rank
Top 15 hybrids based on PC1BI							Top 15 hybrids based on PC2BI							Top 15 hybrids based on direct LD_AV					
E154	4.8	1.7	2.5	6.5	-8.7	1	E28	5.1	1.4	1.1	5.0	1.4	1	E163	4.6	1.7	2.5	5.1	1
E163	4.6	1.7	2.5	5.1	-9.3	2	E155	4.8	1.6	1.7	4.1	0.3	2	E152	4.7	1.6	3.8	4.4	2
E133	4.9	2.1	7.2	7.8	-9.4	3	E156	4.9	1.7	2.4	4.5	-0.3	3	E146	4.7	1.6	6.1	4.1	3
E116	4.9	1.7	2.9	6.1	-9.5	4	E154	4.8	1.7	2.5	6.5	-0.3	4	E164	4.7	1.7	3.7	4.3	4
E141	4.7	2.2	7.9	7.6	-9.5	5	E163	4.6	1.7	2.5	5.1	-0.5	5	E165	4.7	1.9	9.2	4.7	5
E78	4.9	1.9	5.6	7.0	-9.6	6	E12	5.3	1.6	3.2	5.3	-0.7	6	E141	4.7	2.2	7.9	7.6	6
E159	5.0	1.9	7.0	7.7	-9.7	7	E161	4.9	1.7	2.9	5.7	-0.7	7	E183	4.7	1.8	5.2	5.7	7
E34	4.9	1.5	4.1	6.2	-9.8	8	E116	4.9	1.7	2.9	6.1	-0.8	8	E169	4.7	2.1	6.6	5.2	8
E148	4.8	1.8	4.1	5.8	-9.8	9	E1	5.2	1.7	3.3	4.7	-1.0	9	E148	4.8	1.8	4.1	5.8	9
E74	5.2	1.6	3.8	6.8	-9.8	10	E172	5.3	1.6	3.5	4.1	-1.0	10	E139	4.8	2.0	7.8	4.7	10
E161	4.9	1.7	2.9	5.7	-9.9	11	E74	5.2	1.6	3.8	6.8	-1.2	11	E185	4.8	1.9	7.3	5.7	11
E115	5.0	1.8	7.0	7.4	-10.0	12	E126	5.1	1.8	3.4	4.0	-1.2	12	E89	4.8	2.2	11.4	4.2	12
E181	4.8	2.1	7.5	7.0	-10.2	13	E18	4.9	1.7	3.5	5.0	-1.3	13	E117	4.8	2.3	12.8	5.7	13
E111	5.1	1.7	5.3	6.6	-10.3	14	E38	5.3	1.7	3.7	3.0	-1.4	14	E170	4.8	2.0	4.6	4.9	14
E183	4.7	1.8	5.2	5.7	-10.3	15	E130	5.2	1.7	3.8	5.9	-1.5	15	E155	4.8	1.6	1.7	4.1	15
Bottom 5 Hybrids							Bottom 5 Hybrids							Bottom 5 Hybrids					
E189	5.6	2.6	20.0	5.1	-19.3	188	E140	5.1	2.5	19.2	5.7	-16.7	188	E41	5.6	2.0	3.9	3.4	188
E188	5.0	2.5	20.5	3.5	-19.3	189	E189	5.6	2.6	20.0	5.1	-17.2	189	E189	5.6	2.6	20	5.1	189
E127	5.0	2.9	23.9	2.8	-21.5	190	E188	5.0	2.5	20.5	3.5	-18.1	190	E2	5.6	1.8	5.9	2.7	190
E192	6.1	2.3	17.4	3.1	-21.7	191	E127	5.0	2.9	23.9	2.8	-21.7	191	E192	6.1	2.3	17.4	3.1	191
E191	6.2	2.7	31.3	4.4	-26.3	192	E191	6.2	2.7	31.3	4.4	-27.6	192	E191	6.2	2.7	31.3	4.4	192

LD1, LD2, LD3, Leaf feeding damage score at 7, 14, 21 days after infestation, respectively; LD_AV, Average leaf feeding damage score; ED, Ear damage; ER, Ear rot; GY, Grain yield.

10.3389/fpls.2025.1544010

10.3389/fpls.2025.1544010

TABLE 7 Percentage relative selection gain of each of the 15 best-selected hybrids against the commercial checks WH401 and WH505 under different selection schemes.

PC1BI selection index					PC2BI selection index					Direct LD_AV selection scheme				
Against WH401														
Hybrids	LD_AV	ED	ER	GY	Hybrids	LD_AV	ED	ER	GY	Hybrids	LD_AV	ED	ER	GY
E154	-20.3	-26.6	-85.9	112.0	E28	-15.6	-39.9	-93.5	62.1	E163	-24.5	-26.9	-85.9	67.3
E163	-24.5	-26.9	-85.9	67.3	E155	-21.1	-29.7	-90.0	34.9	E152	-23.1	-29.6	-78.1	43.0
E133	-19.7	-9.10	-58.8	154.5	E156	-19.3	-26.1	-86.2	47.2	E146	-22.8	-30.7	-64.9	34.1
E116	-19.3	-23.9	-83.1	98.1	E154	-20.3	-26.6	-85.9	112.0	E164	-22.6	-23.6	-78.7	39.5
E141	-22.2	-4.70	-54.7	147.6	E163	-24.5	-26.9	-85.9	67.3	E165	-22.3	-16.8	-47.5	52.9
E78	-20.0	-18.2	-68.1	127.2	E12	-13.6	-29.8	-81.8	71.6	E141	-22.2	-4.7	-54.7	147.6
E159	-18.1	-16.6	-60.1	152.7	E161	-18.8	-24.1	-83.1	86.6	E183	-22.2	-22.1	-70.0	85.3
E34	-19.0	-33.7	-76.8	102.7	E116	-19.3	-23.9	-83.1	98.1	E169	-22.1	-7.8	-62.3	69.3
E148	-21.9	-21.4	-76.3	88.4	E1	-14.8	-25.5	-81.4	54.4	E148	-21.9	-21.4	-76.3	88.4
E74	-15.2	-30.5	-78.1	121.4	E172	-13.0	-31.6	-79.9	35.5	E139	-21.7	-14.1	-55.5	53.6
E161	-18.8	-24.1	-83.1	86.6	E74	-15.2	-30.5	-78.1	121.4	E185	-21.6	-15.3	-58.4	84.8
E115	-18.1	-22.6	-59.7	140.6	E126	-16.4	-21.9	-80.3	30.9	E89	-21.5	-5.3	-34.4	38.3
E181	-20.6	-5.8	-56.9	127.6	E18	-18.6	-23.7	-80.2	64.6	E117	-21.4	-0.4	-26.5	86.5
E111	-16.7	-25.6	-69.6	117.2	E38	-12.5	-24.1	-78.7	-1.8	E170	-21.4	-13.3	-73.6	60.3
E183	-22.2	-22.1	-70.0	85.3	E130	-15.0	-23.0	-78.5	91.3	E155	-21.1	-29.7	-90.0	34.9
Against WH501														
E154	-22.4	-38.9	-92.1	47.1	E28	-17.8	-50.0	-96.4	12.4	E163	-26.4	-39.2	-92.1	16
E163	-26.4	-39.2	-92.1	16.0	E155	-23.1	-41.5	-94.4	-6.4	E152	-25.1	-41.5	-87.8	-0.8
E133	-21.8	-24.4	-77.0	76.6	E156	-21.4	-38.5	-92.3	2.1	E146	-24.8	-42.3	-80.4	-7.0
E116	-21.4	-36.7	-90.6	37.4	E154	-22.4	-38.9	-92.1	47.1	E164	-24.7	-36.4	-88.1	-3.2
E141	-24.2	-20.8	-74.7	71.8	E163	-26.4	-39.2	-92.1	16.0	E165	-24.3	-30.8	-70.7	6.1
E78	-22.0	-31.9	-82.2	57.6	E12	-15.8	-41.6	-89.8	19.0	E141	-24.2	-20.8	-74.7	71.8
E159	-20.2	-30.6	-77.8	75.3	E161	-20.9	-36.9	-90.6	29.5	E183	-24.2	-35.2	-83.3	28.6
E34	-21.2	-44.8	-87.1	40.6	E116	-21.4	-36.7	-90.6	37.4	E169	-24.1	-23.3	-79.0	17.4
E148	-23.9	-34.6	-86.8	30.7	E1	-17.1	-38.0	-89.6	7.1	E148	-23.9	-34.6	-86.8	30.7
E74	-17.4	-42.2	-87.8	53.6	E172	-15.2	-43.1	-88.8	-6.0	E139	-23.7	-28.5	-75.2	6.6
E161	-20.9	-36.9	-90.6	29.5	E74	-17.4	-42.2	-87.8	53.6	E185	-23.6	-29.6	-76.8	28.2
E115	-20.3	-35.6	-77.5	66.9	E126	-18.6	-35.0	-89.0	-9.2	E89	-23.5	-21.3	-63.5	-4.1
E181	-22.7	-21.6	-76.0	57.9	E18	-20.8	-36.6	-89.0	14.2	E117	-23.4	-17.2	-59.0	29.4
E111	-18.9	-38.1	-83.1	50.7	E38	-14.8	-36.9	-88.1	-31.9	E170	-23.4	-27.9	-85.3	11.2
E183	-24.2	-35.2	-83.3	28.6	E130	-17.2	-35.9	-88.0	32.7	E155	-23.1	-41.5	-94.4	-6.4

LD1, LD2, LD3, Leaf feeding damage score at 7, 14, 21 days after infestation, respectively; ED, Ear damage; ER, Ear rot; GY, Grain yield; PC1BI, Principal component 1-based index; PC2BI, Principal component 2-based index.

the $7tha^{-1}$ threshold. These genotypes warrant further testing as they may carry additional traits beneficial for a breeding program (Ambrósio et al., 2024). The combined selection performance of the two indices (PC1BI and PC2BI), highlighted overlapping selections among the top 15 hybrids. Several of these hybrids were also selected through a direct selection scheme based on LD_AV. Hybrid E163 emerged as a promising genotype, as it was selected across all the selection schemes. The promising genotypes are likely to have a higher frequency of stable and adaptable alleles for both GY and FAW resistance which could be valuable for breeding

10.3389/fpls.2025.1544010

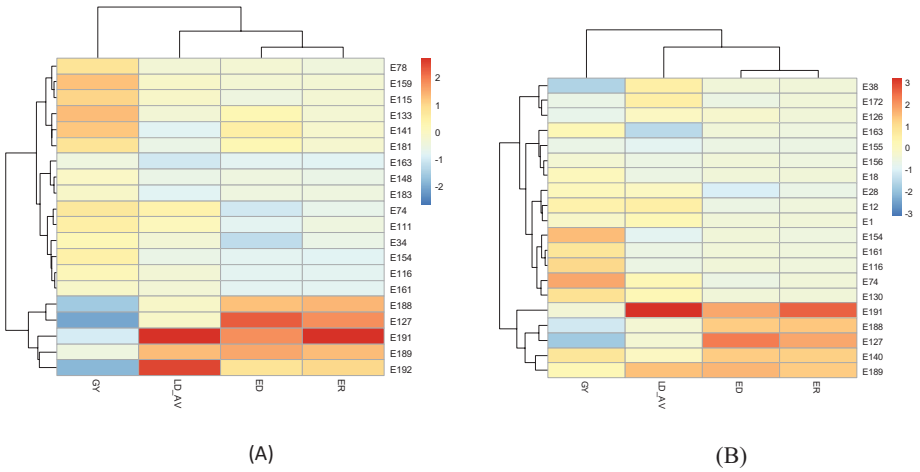


FIGURE 4
Heatmaps displaying relationships between FAW resistance parameters and genotypes based on top 15 and bottom 5 hybrids selected using: **(A)** PC1BI and **(B)** PC2BI-selected hybrids. On the scale, deep red color was desirable for traits that were desired for increase, while deep blue was required for traits that were desired for decrease.

purposes. Genotypes commonly selected across indices exhibit extensive adaptability alleles and perform well across a variety of environments (Zuffo et al., 2020; Ambrósio et al., 2024). This study was carried out under controlled environments, which excluded other traits important for selection of adapted high yielding maize hybrids, for example response to foliar diseases. Additional field studies across a variety of stress conditions often encountered by a large proportion of smallholder farmers are required to validate the yield stability and disease resistance of the selected hybrids for broader adaptation.

Potential future application of PC-based indices

Presently, there are no published reports on the implementation of selection indices to select FAW resistant genotypes in tropical maize in SSA where the pest is relatively new. The PCA approach used in this study offers a framework for the construction and use of an economic weight-free multi-trait selection index for the identification of FAW-resistant genotypes in SSA. This would overcome the limits paused by the need to determine economic weights which can limit gains from selection. However, given the multiple stresses present in SSA (e.g. Makumbi et al., 2018), the development of multiple stress-tolerant varieties suitable for farmers requires selection for a wide range of traits. Therefore, applying a PC-based selection index that includes other important traits in SSA is crucial for optimizing selection gains across multiple traits. Such traits include but are not limited to agronomic traits like lodging resistance, plant height, ear rot, poor husk cover, fungal and viral diseases. Incorporating many traits will help the validation of the PC-based indices as a handy tool for breeders to select for multiple traits. A further area of study could be the incorporation of farmer responses in the PC-based indices. Data collection on a large number of agronomic and disease traits is time consuming and errors can be introduced. We propose integrating advancement in technology during data collection as breeders validate and test indices. For instance, high-throughput phenotyping methods (e.g., Araus et al., 2018) could be used to effectively collect real-time and high-quality data on FAW leaf feeding and ear damage, which are typically costly and labor-intensive traits to measure at multiples intervals. Use of high-throughput phenotyping methods would help to reduce labor costs for breeding programs while increasing efficiency and precision in data collection. Similar high-throughput approaches could be implemented for other agronomic traits like plant and ear height, ear and kernel traits, foliar leaf diseases, and virus diseases like maize lethal necrosis and maize streak virus to further reduce the cost of breeding operations. With reduction in cost, the resources can then be allocated to other breeding operations like acquisition of electronic data collection equipment that improve efficiency and drive genetic gain.

Conclusions

Two principal component (PC)-based indices were developed and evaluated for their effectiveness in selection of FAW-resistant hybrids and achieving desired gains in FAW resistance breeding. Both PC indices explained most of the variation among the genotypes, with PC1BI accounting for most of the variation for FAW leaf feeding damage and grain yield. Larger desired gains were achieved using PC1BI. The PC-based indices resulted in more balanced multi-trait genetic gains compared to the single-trait selection approach. The use of PC-based index selection presents a promising, economic weight-free selection tool to maximize genetic gains in FAW resistance breeding programs. However, it should also incorporate additional desirable agronomic and adaptive traits beyond the FAW-related parameters.

Data availability statement

The data sets used in this study have all been analyzed and summarized in the main Tables, Figures, and Supplementary Tables presented.

Author contributions

WW: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing. DM: Conceptualization, Formal analysis, Investigation, Methodology, Supervision, Validation, Writing – review & editing, Writing – original draft. GA: Investigation, Supervision, Writing – review & editing. HZ: Investigation, Supervision, Writing – review & editing. AB: Investigation, Writing – review & editing. MW: Supervision, Writing – review & editing. DK: Writing – review & editing. BP: Funding acquisition, Project administration, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. Funding for the research was provided by the International Maize and Wheat Improvement Center (CIMMYT) under the USAID-FAW project, and by the Bill and Melinda Gates Foundation (BMGF) through the Accelerating Genetic Gains in Maize and Wheat for Improved Livelihoods project (INV–003439). The first author was supported by the Eastern and Southern Africa Higher Education Center of Excellence Project: IDA Credit 5794-ET, under the Africa Centre of Excellence in Climate Smart Agriculture and Biodiversity Conservation for a PhD study. Additional support from USAID Feed the Future (USAID-Ftf) Uganda Agricultural Research Activity is acknowledged.

Conflict of interest

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

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.2025.1544010/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Scree plot showing eigenvalues for the seven PCs.

References

Agudelo-Gómez, D. A., Agudelo-Trujillo, J. H., and Cerón-Muñoz, M. F. (2016). Selection index for meat and milk traits of buffaloes in Colombia. *Livestock Sci.* 191, 6–11. doi: 10.1016/j.livsci.2016.06.016

Alvarado, G., Rodríguez, F. M., Pacheco, A., Burgueño, J., Crossa, J., Vargas, M., et al. (2020). META-R: A software to analyze data from multi-environment plant breeding trials. *Crop J.* 8, 745–756. doi: 10.1016/j.cj.2020.03.010

Ambrósio, M., Daher, R. F., Santos, R. M., Santana, J. G. S., Vidal, A. K. F., Nascimento, M. R., et al. (2024). Multi-trait index: selection and recommendation of superior black bean genotypes as new improved varieties. *BMC Plant Biol.* 24, 1–12. doi: 10.1186/s12870-024-05248-5

Anshori, M. F., Musa, Y., Dungga, N. E., Widiayani, N., Arifin, A. S., Masniawati, A., et al. (2024). A new approach for selection of transgressive segregants in F3 populations based on selection index and anthocyanin content in cayenne pepper. *Front. Sustain. Food Syst.* 8. doi: 10.3389/fsufs.2024.1288579

Anshori, M. F., Purwoko, B. S., Dewi, I. S., Suwarno, W. B., and Ardie, S. W. (2022). Salinity tolerance selection of doubled-haploid rice lines based on selection index and factor analysis. *AIMS Agric. Food.* 7, 520–535. doi: 10.3934/agrfood.2022032

Anyanda, G. N., Bruce, A. Y., Makumbi, D., Ahonsi, M., Kahuthia-Gathu, R., Namikoye, S. E., et al. (2022). Reproductive potential of fall armyworm *Spodoptera frugiperda* (JE Smith) and effects of feeding on diverse maize genotypes under artificial infestation. *Front. Insect Sci.* 2. doi: 10.3389/finsc.2022.950815

Araus, J. L., Kefauver, S. C., Zaman-Allah, M., Olsen, M. S., and Cairns, J. E. (2018). Translating high-throughput phenotyping into genetic gain. *Trends Plant Sci.* 23, 451–466. doi: 10.1016/j.tplants.2018.02.001

Asfaw, D. M., Asnakew, Y. W., Sendkie, F. B., Abdulkadr, A. A., Mekonnen, B. A., Tiruneh, H. D., et al. (2024). Analysis of constraints and opportunities in maize production and marketing in Ethiopia. *Heliyon* 10, 1–7. doi: 10.1016/j.heliyon.2024.e39606

Badu-Apraku, B., and Fakorede, M. A. B. (2017). Maize in Sub-Saharan Africa: Importance and Production Constraints. In *Advances in Genetic Enhancement of Early and Extra-Early Maize for Sub-Saharan Africa*. (Springer International Publishing), 3–10. doi: 10.1007/978-3-319-64852-1_1

Batista, L. G., Gaynor, R. C., Margarido, G. R. A., Byrne, T., Amer, P., Gorjanc, G., et al. (2021). Long-term comparison between index selection and optimal independent culling in plant breeding programs with genomic prediction. *PloS One* 16, e0235554. doi: 10.1371/journal.pone.0235554

Cerón-Rojas, J., and Crossa, J. (2022). The statistical theory of linear selection indices from phenotypic to genomic selection. *Crop Sci.* 62, 537–563. doi: 10.1002/csc2.20676

Crispim-Filho, A. J., dos Santos, F. P., Pinto, J. F. N., Melo, P. G. S., dos Reis, E. F., and Mendes-Resende, M. P. (2020). Dealing with multiple traits in maize: A new approach for selecting progenies. *Crop Sci.* 60, 3151–3165. doi: 10.1002/csc2.20292

Davis, F. M., and Williams, W. P. (1992). *Visual rating scales for screening whorl-stage corn for resistance to fall armyworm*. Mississippi Agricultural & Forestry Experiment Station, Technical Bulletin 186, Mississippi, State University, MS 39762, USA.

Day, R., Abrahams, P., Bateman, M., Beale, T., Clottey, V., Cock, M., et al. (2017). Fall armyworm: Impacts and implications for Africa. *Outlooks Pest Manage.* 28, 196–201. doi: 10.1564/v28_oct_02

Ekpa, O., Palacios-Rojas, N., Kruseman, G., Fogliano, V., and Linnemann, A. R. (2018). Sub- Saharan African maize-based foods: Technological perspectives to increase the food and nutrition security impacts of maize breeding programmes. *Global Food Secur.* 17, 48–56. doi: 10.1016/j.gfs.2018.03.007

Ekpa, O., Palacios-Rojas, N., Kruseman, G., Fogliano, V., and Linnemann, A. R. (2019). Sub- Saharan African maize-based foods - processing practices, challenges and opportunities. *Food Rev. Int.* 35, 609–639. doi: 10.1080/87559129.2019.1588290

Erenstein, O., Jaleta, M., Sonder, K., Mottaleb, K., and Prasanna, B. M. (2022). Global maize production, consumption and trade: trends and R&D implications. *Food Secur.* 14, 1295–1319. doi: 10.1007/s12571-022-01288-7

Falconer, D. S., and Mackay, T. F. C. (1996). *Introduction to quantitative genetics. 4th ed* (Essex, UK: Longman Group Ltd.).

FAO (2017). *The future of food and agriculture – Trends and challenges* (Rome: Food and Agriculture Organization of the United Nations). Available online at: http://www.fao.org/3/a-i6583e.pdf (Accessed November 15, 2024).

FAO (2021). *Evaluation of FAO's contributions to Sustainable Development Goal 2 - Control of transboundary plant diseases and pests* (Rome: Food and Agriculture Organization of the United Nations).

Farid, M., Anshori, M. F., and Ridwan, I. (2022). Tomato F3 lines development and its selection index based on narrow-sense heritability and factor analysis. *Biodiversitas* 23, 5790–5797. doi: 10.13057/biodiv/d231132

Goergen, G., Lava, K. P., Sankung, S. B., Togola, A., and Tamò, M. (2016). First report of outbreaks of the fall armyworm, *spodoptera frugiperda* (J.E. Smith) (Lepidoptera, noctuidae): A new alien invasive pest in west and central Africa. *PloS One*. 11(10), e0165632. doi: 10.1186/s12870-024-05248-5

Hallauer, A. R., Carena, M., and Miranda Filho, J. B. (2010). *Quantitative genetics in maize breeding* (New York: Springer).

Iezzoni, A. F., and Pritts, M. P. (1991). Applications of principal component analysis to horticultural research. *HortScience* 26, 334–338. doi: 10.21273/HORTSCI.26.4.334

Kamweru, I., Anani, B. Y., Beyene, Y., Makumbi, D., Prasanna, B. M., and Gowda, M. (2022). Genomic analysis of resistance to fall armyworm (*Spodoptera frugiperda*) in CIMMYT maize lines. *Genes* 13, 1–19. doi: 10.3390/genes13020251

Kamweru, I., Beyene, Y., Anani, B., Adetimirin, V. O., Prasanna, B. M., and Gowda, M. (2023). Hybrid breeding for fall armyworm resistance: Combining ability and hybrid prediction. *Plant Breed.* 142, 607–620. doi: 10.1111/pbr.13129

Kasoma, C., Shimelis., H., Laing, D. M., Shayanowako., A., and Mathew, I. (2021). Combining ability of maize genotypes for fall armyworm (*Spodoptera frugiperda* J.E. Smith) resistance, yield and yield-related traits. *Crop Prot.* 149, 1–9. doi: 10.1016/j.cropro.2021.105762

Kassie, M., Wossen, T., De Groote, H., Tefera, T., Sevgan, S., and Solomon Balew, S. (2020). Economic impacts of fall armyworm and its management strategies: evidence from southern Ethiopia. *Eur. Rev. Agric. Economics* 47, 1473–1501. doi: 10.1093/erae/jbz048

Kempthorne, O., and Nordskog, A. W. (1959). Restricted selection indices. *Biometrics* 15, 10–19. doi: 10.2307/2527598

Kolde, R. (2019). *pheatmap: Pretty Heatmaps version version 1.0.12*. doi: 10.32614/CRAN.package.pheatmap

Logrieco, A. F., Battilani, P., Camardo Leggieri, M., Haesaert, G., Jing, X., Lanubile, A., et al. (2021). Perspectives on global mycotoxin issues and management from MycoKey Maize working group. *Plant Dis.* 105, 525–537. doi: 10.1094/PDIS-06-20-1322-FE

Lopez-Cruz, M., and de los Campos, G. (2021). Optimal breeding-value prediction using a sparse selection index. *Genetics* 218, iyab030. doi: 10.1093/GENETICS/IYAB030

Makumbi, D., Assanga., S., Diallo., A., Magorokosho., C., Asea., G., Worku., M., et al. (2018). Genetic analysis of tropical midaltitude- adapted maize populations under stress and nonstress conditions. *Crop Sci.* 58, 1492–1505. doi: 10.2135/cropsci2017.09.0531

Matias, F. I., do Valle, C. B., Gouveia, B. T., Moro, G. V., and Barrios, S. C. L. (2020). Using additive indices and principal components to select sexual genitors and hybrids of *Urochloa decumbens*. *Crop Breed. Appl. Biotechnol.* 20, 1–8. doi: 10.1590/1984-70332020v20n2a18

Mhike, X., Patrick Okori., P., Magorokosho., C., and Ndlela., T. (2012). Validation of the use of secondary traits and selection indices for drought tolerance in tropical maize (*Zea mays* L.). *Afr. J. Plant Sci.* 6, 96–102. doi: 10.5897/AJPS11.179

Mihm, J. A. (1997). *Insect Resistant Maize: Recent Advances and Utilization; Proceedings of an International Symposium held at the International Maize and Wheat Improvement Center (CIMMYT) 27 November-3December 1994* (Mexico, D.F: CIMMYT). Available at: http://hdl.handle.net/10883/744 (Accessed November 15, 2024).

Montezano, D. G., Specht, A., Sosa-Gómez, D. R., Roque-Specht, V. F., Sousa-Silva, J. C., de-Paula-Moraes, S. V., et al. (2018). Host plants of *Spodoptera frugiperda* (Lepidoptera: Noctuidae) in the Americas. *Afr. Entomol.* 26, 286–300. doi: 10.4001/003.026.0286

Norman, P. E., Agre, P. A., Asiedu, R., and Asfaw, A. (2022). Multiple-traits selection in white Guinea yam (*Dioscorea rotundata*) genotypes. *Plants* 11, 3003. doi: 10.3390/plants11213003

Nyquist, W. E. (1991). Estimation of heritability of prediction of response in plant populations. *Crit. Rev. Plant Sci.* 10, 235–322. doi: 10.1080/07352689109382313

Olivoto, T., and Nardino, M. (2021). MGIDI: towards an effective multivariate selection in biological experiments. *Bioinformatics* 37, 1383–1389. doi: 10.1093/bioinformatics/btaa981

Oloyede-Kamiyo, Q. O. (2019). Efficiency of index-based selection methods for stem borer resistance in maize (*Zea mays* L.). *J. Crop Sci. Biotech.* 22, 205–211. doi: 10.1007/s12892-017-0130-0

Otim, M. H., Fiaboe, K. K. M., Akello, J., Mudde, B., Obonyom, A. T., Bruce, A. Y., et al. (2021). Managing a Transboundary Pest: The Fall Armyworm on Maize in Africa. In *Moths and Caterpillars* (IntechOpen). doi: 10.5772/intechopen.96637

Otim, M. H., Tay, W. T., Walsh, T. K., Kanyesigye, D., Adumo, S., Abongosi, J., et al. (2018). Detection of sister-species in invasive populations of the fall armyworm *Spodoptera frugiperda* (Lepidoptera: Noctuidae) from Uganda. *PloS One* 13, 1–18. doi: 10.1371/journal.pone.0194571

Padjung, R., Farid, M., Musa, Y., Anshori, M. F., Nur, A., and Masnenong, A. (2021). Drought-adapted maize line based on morphophysiological selection index. *Biodiversitas* 22, 4028–4035. doi: 10.13057/biodiv/d220951

Palumbo, R., GonÇAlves, A., Gkrillas, A., Logrieco, A., Dorne, J., Dall'Asta, C., et al. (2020). Mycotoxins in maize: mitigation actions, with a chain management approach. *Phytopathol. Mediterr.* 59, 5–28. doi: 10.14601/Phyto-11142

Prasanna, M. B., Bruce, A., Beyene, Y., Makumbi, D., Gowda, M., Asim, M., et al. (2022). Host plant resistance for fall armyworm management in maize: relevance, status and prospects in Africa and Asia. *Theor. Appl. Genet.* 135, 3897–3916. doi: 10.1007/s00122-022-04073-4

Prasanna, B. M., Bruce, A., Winter, S., Otim, M., Asea, G., Sevgan, S., et al. (2018). "Host plant resistance to fall armyworm," in *Fall armyworm in Africa: a guide for integrated pest management, 1st edn*. Ed. B. M. Prasanna, et al (CDMX, CIMMYT, Mexico), 45–62.

Prasanna, B. M., Cairns, J. E., Zaidi, P. H., Beyene, Y., Makumbi, D., Gowda, M., et al. (2021). Beat the stress: breeding for climate resilience in maize for the tropical rainfed environments. *Theor. Appl. Genet.* 134, 1729–1752. doi: 10.1007/s00122-021-03773-7

Pruter, L. S., Weaver, M., and Brewer, M. J. (2020). Overview of risk factors and strategies for management of Insect-Derived Ear Injury and Aflatoxin Accumulation for Maize Grown in Subtropical Areas of North America. *J. Integrated Pest Management* 11, 1–8. doi: 10.1093/jipm/pmaa005

Ramirez-Cabral, N. Y. Z., Kumar, L., and Shabani, F. (2017). Future climate scenarios project a decrease in the risk of fall armyworm outbreaks. *J. Agric. Sci.* 155, 1219–1238. doi: 10.1017/S0021859617000314

Santos, I. G., Cruz, C. D., Nascimento, M., Rosado, R. D. S., and Ferreira, R. P. D. (2018). Direct, indirect and simultaneous selection as strategies for alfalfa breeding on forage yield and nutritive value. *Pesqui. Agropecu. Trop.* 48, 178–189. doi: 10.1590/1983-40632018v4851950

Schneider, L., Rebetez, M., and Rasmann, S. (2022). The effect of climate change on invasive crop pests across biomes. *Curr. Opin. Insect Sci.* 50, 1–5. doi: 10.1016/j.cois.2022.100895

Singh, B. U., Rao, K. V., and Sharma, H. C. (2011). Comparison of selection indices to identify sorghum genotypes resistant to the spotted stemborer *Chilo partellus* (Lepidoptera: Noctuidae). *Int. J. Trop. Insect Sci.* 31, 38–51. doi: 10.1017/S1742758411000105

Sisay, B., Simiyu, J., Mendesil, E., Likhayo, P., Ayalew, G., Mohamed, S., et al. (2019). Fall armyworm, *spodoptera frugiperda* infestations in east Africa: assessment of damage and parasitism. *Insects* 10, 1–10. doi: 10.3390/insects10070195

Skendžić, S., Zovko, M., Živković, I. P., Lešić, V., and Lemić, D. (2021). The impact of climate change on agricultural insect pests. *Insects* 12, 1–32. doi: 10.3390/insects12050440

Smith, H. F. (1936). A discriminant function for plant selection. *Ann. Eugenics* 7, 240–250. doi: 10.1111/j.1469-1809.1936.tb02143.x

Smith, O. S., Hallauer, A. R., and Russell, W. A. (1981). Use of index selection in recurrent selection programs in maize. *Euphytica* 30, 611–618. doi: 10.1007/BF00038788

Soujanya, P. L., Sekhar, J. C., Yathish, K. R., Karjagi, C. G., Rao, K. S., Suby, S. B., et al. (2022). Leaf damage based phenotyping technique and its validation against fall armyworm, *spodoptera frugiperda* (J. E. Smith), in maize. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.906207

Stromberg, L. D., and Compton, W. A. (1989). Ten cycles of full-sib selection in maize. *Crop Sci.* 29, 1170–117. doi: 10.2135/cropsci1989.0011183X002900050014x

Subandi,, Compton, W. A., and Empig, L. T. (1973). Comparison of the efficiencies of selection indices for three traits in two variety crosses of corn. *Crop Sci.* 13, 184–186. doi: 10.2135/cropsci1973.0011183X001300020011x

Tai, G. C. C. (1977). Index selection with desired gains. *Crop Sci.* 17, 182–183.

Tambo, J. A., Kansiime, M. K., Mugambi, I., Rwomushana, I., Kenis, M., Day, R. K., et al. (2020). Understanding smallholders' responses to fall armyworm (*Spodoptera frugiperda*) invasion: Evidence from five African countries. *Sci. Total Environment* 740, 1–11. doi: 10.1016/j.scitotenv.2020.140015

Tang, H., Xu, C., Jiang, Y., Wang, J., Wang, Z., and Tian, L. (2021). Evaluation of physical characteristics of typical maize seeds in a cold area of north China based on principal component analysis. *Processes* 9, 1–1. doi: 10.3390/pr9071167

Uganda Bureau of Statistics (UBOS) (2022). Statistical abstract. Available online at: http://www.ubos.org/onlinefiles/uploads/ubos/pdfdocuments/abstracts/StatisticalAbstract2013.pdf (Accessed October 8, 2024).

Warburton, M. L., Woolfolk, S. W., Smith, J. S., Hawkins, L. K., Castano-Duque, L., Lebar, M. D., et al. (2023). Genes and genetic mechanisms contributing to fall armyworm resistance in maize. *Plant Genome* 16, 1–13. doi: 10.1002/tpg2.20311

Widstrom, N. W., Williams, W. P., Wiseman, B. R., and Davis, F. M. (1992). Recurrent selection for resistance to leaf feeding by fall armyworm on maize. *Crop Sci.* 32, 1171–1174. doi: 10.2135/cropsci1992.0011183X003200050022x

Widstrom, N. W., Wiseman, B. R., and McMillan, W. W. (1982). Response to index selection in maize for resistance to ear damage by the corn earworm. *Crop Sci.* 22, 843–846. doi: 10.2135/cropsci1982.0011183x002200040034x

William, R. (2024). *psych: procedures for psychological, psychometric, and personality research* (Evanston, Illinois: Northwestern University). Available at: https://CRAN.R-project.org/package=psych (Accessed November 21, 2024).

Williams, J. S. (1962). The evaluation of a selection index. *Biometrics* 18, 375–393. doi: 10.2307/2527479

Yan, X.-R., Wang, Z.-Y., Feng, S.-Q., Zhao, Z.-H., and Li, Z.-H. (2022). Impact of Temperature Change on the Fall Armyworm, *Spodoptera frugiperda* under Global Climate Change. *Insects* 13, 1–17. doi: 10.3390/insects13110981

Zuffo, A. M., Steiner, F., Aguilera, J. G., Teodoro, P. E., Teodoro, L. P. R., and Busch., A. (2020). Multi-trait stability index: A tool for simultaneous selection of soya bean genotypes in drought and saline stress. *J. Agron. Crop Sci.* 206, 815–822. doi: 10.1111/jac.12409

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 31 March 2025
DOI 10.3389/fpls.2025.1536981

Check for updates

OPEN ACCESS

EDITED BY
Mahendar Thudi,
Fort Valley State University, United States

REVIEWED BY
Zelalem Mersha,
Virginia State University, United States
Hari Kishan Sudini,
International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), India

*CORRESPONDENCE
Manje Gowda
m.gowda@cgair.org

RECEIVED 29 November 2024
ACCEPTED 07 March 2025
PUBLISHED 31 March 2025

CITATION
Suresh LM, Gowda M, Beyene Y, Makumbi D, Manigben KA, Burgueño J, Okayo R, Woyengo VW and Prasanna BM (2025) Identification of gray leaf spot–resistant donor lines in tropical maize germplasm and their agronomic performance under artificial inoculation.
Front. Plant Sci. 16:1536981.
doi: 10.3389/fpls.2025.1536981

COPYRIGHT
© 2025 Suresh, Gowda, Beyene, Makumbi, Manigben, Burgueño, Okayo, Woyengo and Prasanna. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Identification of gray leaf spot–resistant donor lines in tropical maize germplasm and their agronomic performance under artificial inoculation

L. M. Suresh[1], Manje Gowda[1]*, Yoseph Beyene[1], Dan Makumbi[1], Kulai Amadu Manigben[1,2,3], Juan Burgueño[4], Robert Okayo[1], Vincent W. Woyengo[5] and Boddupalli M. Prasanna[1]

[1]Global Maize Program, International Maize and Wheat Improvement Center (CIMMYT), Nairobi, Kenya, [2]West Africa Centre for Crop Improvement (WACCI), University of Ghana, Accra, Ghana, [3]Maize Improvement Program, Council for Scientific and Industrial Research (CSIR)-Savanna Agricultural Research Institute, Nyankpala, Ghana, [4]Biometrics and Statistics Unit, International Maize and Wheat Improvement Center (CIMMYT), Texcoco, Edo de México, Mexico, [5]Kenya Agricultural and Livestock Research Organization, Kakamega Research Institute, Kakamega, Kenya

Gray leaf spot (GLS) disease is caused by two fungal pathogens, *Cercospora zeae-maydis* and *Cercospora zeina*. The current study evaluated 427 elite tropical/subtropical lines for their responses to GLS under artificial inoculation in Kakamega in western Kenya for 4 years. Furthermore, a subset of 140 lines was used for a high-resolution genome-wide association study (GWAS) for GLS resistance. Among the 427 lines evaluated, 14 were identified as resistant on the basis of a <4 (on a scale of 1–9) GLS disease severity score. Among these 14 lines, three lines, namely CML540, CML559, and CML566, are also known for resistance to MSV, tolerance to drought, and resistance to MLN, respectively. The phenotypic evaluation revealed significant ($P < 0.01$) genotypic and genotype x environment interaction variances and moderate to high heritability for GLS disease severity, area under disease progress curve (AUDPC), and other agronomic traits. GLS disease severity traits were negatively and significantly correlated ($P < 0.01$) with anthesis date, silking date, plant height, and ear height. A subset of 140 lines was genotyped with 33,740 DART-GBS SNP markers. Population structure and principal component analysis grouped the lines into two major clusters with moderate structure in the population. GWAS revealed 13 and 11 SNPs significantly associated with GLS disease severity and AUDPC values. Six among the 13 SNPs detected for GLS resistance are overlapped with earlier studies, which can be used for fine mapping and improvement of GLS resistance through marker-assisted selection. However, SNPs on chromosomes 9 and 10 were unique to the present study. Genomic prediction on GLS traits revealed moderate to high prediction correlations, suggesting its usefulness in the selection of desirable candidates with favorable alleles for GLS resistance. Overall, 14 GLS resistance lines identified in this study can be used as donor lines in both genetic studies and resistance breeding programs.

KEYWORDS

maize, gray leaf spot, genome-wide association (GWA) study, genomic prediction, DART genotyping by sequencing

Introduction

Maize is the most important cereal crop in sub-Saharan Africa (SSA), growing over 35 million hectares, with an average production of over 70 million metric tons of grain (FAOSTAT, 2021). The crop is mainly grown by several million smallholder farmers for their food security, income, and livelihoods across SSA (Prasanna et al., 2020, 2021). The average maize yield in SSA is very low (~1.7 tons/hectare) compared to the world average (~5 tons/hectare) due to various abiotic and biotic stresses (Erenstein et al., 2022). Maize in eastern and southern Africa, specifically in highlands, is affected by many fungal diseases, mainly Turcicum leaf blight (TLB), gray leaf spot (GLS), and Fusarium ear rot (Shi et al., 2007; Kibe et al., 2020a; Nyaga et al., 2020; Omondi et al., 2023; Ndlovu et al., 2024). GLS caused by *Cercospora zea maydis* (Ward et al., 1999) and *C. zeina*, *maydis* (Crous et al., 2006; Liu and Xu, 2013) are major threats to maize production in the world (Ward et al., 1999; Katwal et al., 2013; Omondi et al., 2023);. During the 1960s and '70s, the disease was first reported in the USA and later spread worldwide and became a key concern for maize production (Liu et al., 2016) with significant economic yield loss. In SSA, the average yield loss exceeds 70% implicating a significant economic and food security concerns (Vivek et al., 2010; Kinyua et al., 2011; Yigrem and Yohannes, 2019).

Pathogens causing GLS are favored by environmental conditions such as high humidity, moderately high temperature, and extended leaf wetness. Initial symptoms of GLS appear on the lower leaves and are progressively observed in the upper leaves later during the season. The characteristic symptoms of mature GLS lesions are gray to tan in color, sharply rectangular, long and narrow, and run parallel to the leaf veins (Latterell, 1983; Katwal et al., 2013). The GLS damage has been linked to loss of photosynthetic capability and premature plant death (Latterell, 1983; Dodd and McGee, 1989). Due to the widespread impact of the disease in eastern and southern Africa and globally, there is a critical need to adopt effective disease management strategies. The main disease management approaches include the application of fungicides, cultural practices, and, most importantly, host resistance mechanisms (Katwal et al., 2013). Many maize breeding improvement programs in eastern and southern Africa resorted to develop and deploy GLS resistant maize germplasm, as other methods such as the application of fungicides are neither affordable and economical nor environment friendly (Menkir and Ayodele, 2005; Nzuve, 2013; Shi et al., 2014). Hence, breeding for resistant germplasm through conventional methods and integrating advanced molecular tools is the most effective method to control diseases and ensure maize-based food security in SSA.

Finding the resistance source of germplasm and understanding the genetic basis of resistance is critical in managing plant diseases. However, this process is still in progress for GLS resistance. Earlier studies indicated that GLS resistance is a complex trait controlled by multiple genes with small to moderate additive effects (Berger et al., 2014; Benson et al., 2015; Kibe et al., 2020a; Omondi et al., 2023) and strongly influenced by the environment (Clements et al., 2000). Quantitative trait loci (QTL) mapping is an effective tool to understand the genetic basis of complex traits like GLS resistance. Previous QTL mapping studies identified several genomic regions that confer resistance to GLS, which helped to understand its genetic architecture (Zhang et al., 2012, 2017; Berger et al., 2014; Xu et al., 2014; Mammadov et al., 2015; Du et al., 2020; Qiu et al., 2021). QTLs on chromosome bin 5.04 and 5.06–07 were consistently detected in different mapping studies and are of interest to improve GLS resistance (Martins et al., 2019). Meta-QTL analysis revealed that bin 8.08 on chromosome 8 possesses a cluster of QTLs and significant consensus QTLs for GLS, TLB, and southern leaf blight (SLB) with less than 5 cM of confidence interval and also found to be associated with two nucleotide-binding site (NBS) family of R genes (Ali et al., 2013). QTL studies with different populations also revealed QTLs on chromosome bin 8.05/8.06 has been detected for GLS as well as for other foliar diseases like common rust and common smut (Bubeck et al., 1993; Kerns et al., 1999). The consistent genomic regions identified in these studies are needed to focus on improving GLS resistance. However, not many studies were done on eastern and southern African-adapted germplasm.

However, QTL mapping has its limitations, like low mapping resolution due to few recombinations in population development, and each mapping population represents only two alleles. In addition, low relatedness between mapping populations and breeding populations also hampers the translation of the identified QTL into breeding targets. On the contrary, the linkage disequilibrium-based genome-wide association study (GWAS) in a set of mapping panels that represents a broad diversity of the germplasm in breeding programs is a powerful tool for dissecting oligogenic and polygenic traits. GWAS panels with genetically unrelated individuals are expected to accumulate a large number of historical recombination events from the past, which can help to overcome the problems related to the lack of recombination events (Mammadov et al., 2015). In maize, GWAS has been widely used to identify the allelic variants that contribute to improve resistance to many of the maize diseases like maize lethal necrosis (Gowda et al., 2015; Sitonik et al., 2019; Nyaga et al., 2020), GLS (Shi et al., 2014; Kuki et al., 2018; Kibe et al., 2020a; Omondi et al., 2023), sugarcane mosaic virus (Tao et al., 2013), maize streak virus (Nair et al., 2015), common rust (Zheng et al., 2018; Kibe et al., 2020b, Nyaga et al., 2020), Tar spot complex (Mahuku et al., 2016), Fusarium ear rot (Liu et al., 2021), head smut (Wang et al., 2012), and TLB (Poland et al., 2011; Rashid et al., 2020; Ndlovu et al., 2024). Even though several research groups reported GWAS on GLS resistance, the studies on screening a large number of lines to identify the best donor lines and conducting GWAS on lines adapted to SSA are seldom. Therefore, the present research was designed to screen a large number of locally adapted, widely used elite lines and conduct GWAS for GLS resistance in SSA.

GWAS is widely used to find trait-linked markers. The application of trait-linked markers in breeding is limited to large effect QTLs or markers. On the contrary, genomic selection (GS) uses genome-wide markers and captures variations explained by both large-effect and small-effect QTLs or markers, which is effective for complex traits like grain yield and moderately

complex traits like GLS and TLB (Kibe et al., 2020a). In GS, the training population has genotypes with both phenotypic and marker data and is used to establish prediction models. From the marker effects estimated from the training population, the genomic-estimated breeding values (GEBVs) are predicted in a testing population that has only marker data but no phenotypic data (Meuwissen et al., 2001). By doing this, we can phenotype only selected lines under controlled environments to produce reliable data. Empirical research has shown the advantage of GS for accelerating the genetic gains per unit of time over phenotypic selection. In maize, GS has been widely applied for many traits for inbred line prediction (Zhao et al., 2012; Technow et al., 2013; Beyene et al., 2019, 2021; Ndlovu et al., 2022, 2024; Sadessa et al., 2022) and hybrid performance prediction (Guo et al., 2019; Schrag et al., 2019; Li et al., 2020). These findings demonstrate the potential of GS to help in the selection of elite lines for disease resistance. The objectives of the present study are (i) to evaluate 427 elite lines for 4 years under artificially inoculated conditions and identify the best GLS resistance donor lines, (ii) to identify the genomic regions associated with GLS resistance through GWAS, and (iii) to examine the potential of GS in predicting GLS resistance and other agronomical traits.

Materials and methods

A collection of 427 elite inbred lines widely used in eastern and southern Africa subtropical maize breeding programs was assembled. These inbred lines have potential resistance sources to various foliar diseases, including GLS. The information on the pedigree and other details is provided in Supplementary Table S1. All these lines have been planted to test for their response to GLS under field conditions with artificial inoculation at the KALRO (Kenya Agricultural and Livestock Research Organization) Research Station, Kakamega, Kenya (0°16′N and 34°49′E, 1406 masl) for 4 years (2016, 2017, 2018, and 2019). The monthly average rainfall and temperature (minimum and maximum) in this disease screening period are provided in Figure 1. The disease screening location is a hot spot for foliar diseases, including GLS. However, climate change has affected temperature and rainfall patterns in the region, leading to inconsistent disease infection and expression. To have accurate disease severity data, each trial was artificially inoculated to ensure uniform disease expression.

Four hundred and twenty-seven (427) entries were planted in a 5 × 86 alpha (α)–lattice design, randomized and replicated two times each year, by using CIMMYT's field book (Vivek et al., 2010). These inbred lines were planted in one row of four-meter plots with 15 plants per row. The trial was conducted in the main rainy season (April–September). For these trials, two seeds were planted per hill and later thinned to a single plant per hill three weeks after emergence. This was done to ensure a uniform plant density. All standard agronomic practices were applied during the disease screening period.

Samples collection for pathogen isolation and inoculation

Leaves from five to ten maize plants with representative GLS lesions were sampled from 20 to 25 fields in western Kenya. The pathogen *Cercospora zeina* was isolated from infected leaves using the single spore isolation method (Meisel et al., 2009; Kinyua et al., 2011). The infected leaves were chopped into smaller pieces (2–5 mm^2) and surface-sterilized in 5.25% sodium hypochlorite (NaOCl, pure chemical) solution for approximately 30 seconds. The samples were rinsed in sterile distilled water, dried, and plated on potato dextrose agar (PDA) media supplemented with streptomycin sulfate (0.03 g/liter) and incubated at 30°C for five days to allow the pathogen to sporulate in a growth cabinet under a 12h fluorescent light/dark regime. The sporulating fungi were observed under a light microscope, and the hyphal tips of the correct fungi advancing from the colony margins were sub-cultured onto fresh PDA media as part of the culture purification process. A pure culture was stored at −20°C in the KALRO Kakamega Laboratory.

The pathogen from the mother culture was inoculated and multiplied on a susceptible maize host (hybrid PAN4M-19). The conidia were dislodged with a brush and rinsed with 0.01% Tween 20 and the spore concentration was adjusted to a standardized concentration of 4×10^4 spores/ml using a microscope and hemocytometer and applied to all the leaves (V3 stage) of the maize plants with a small brush, and the inoculation was repeated after seven days. During the inoculation period, we walked along the inter-row valley and ensured a uniform inoculum density across the testing population.

Phenotypic evaluation and data analyses

The inbred lines were evaluated for their responses to GLS in four environments. GLS disease severity is typically at its peak between tasseling and physiological maturity; therefore, disease severity data were recorded at the mid-silking, 77 days after planting (1st scoring called GLS 1) and at the hard dough stage, 105 days after planting (second score called GLS 2). Disease severity was rated plot-wise on the ordinal scale of 1 (highly resistant, without any disease symptoms) to 9 (highly susceptible, with necrosis and completely dead plants). Based on the scoring of 1 to 9 scale disease severity data, we divided the genotypes response as a resistant, moderately resistant, moderately susceptible, and completely susceptible group when the scores were 1 to 4, 4.1 to 5, 5.1 to 7, and 7.1 to 9, respectively. The area under the disease progress curve (AUDPC), a quantitative measure of disease intensity with time, was calculated for each plot to provide a measure of the progression of GLS severity. The AUDPC was computed according to the following equation (Shaner and Finney, 1977):

$$\text{AUDPC} = \sum_{i=1}^{n} \left[\frac{y_i + y_{i+1}}{2}\right](t_{i+1} - t_i)$$

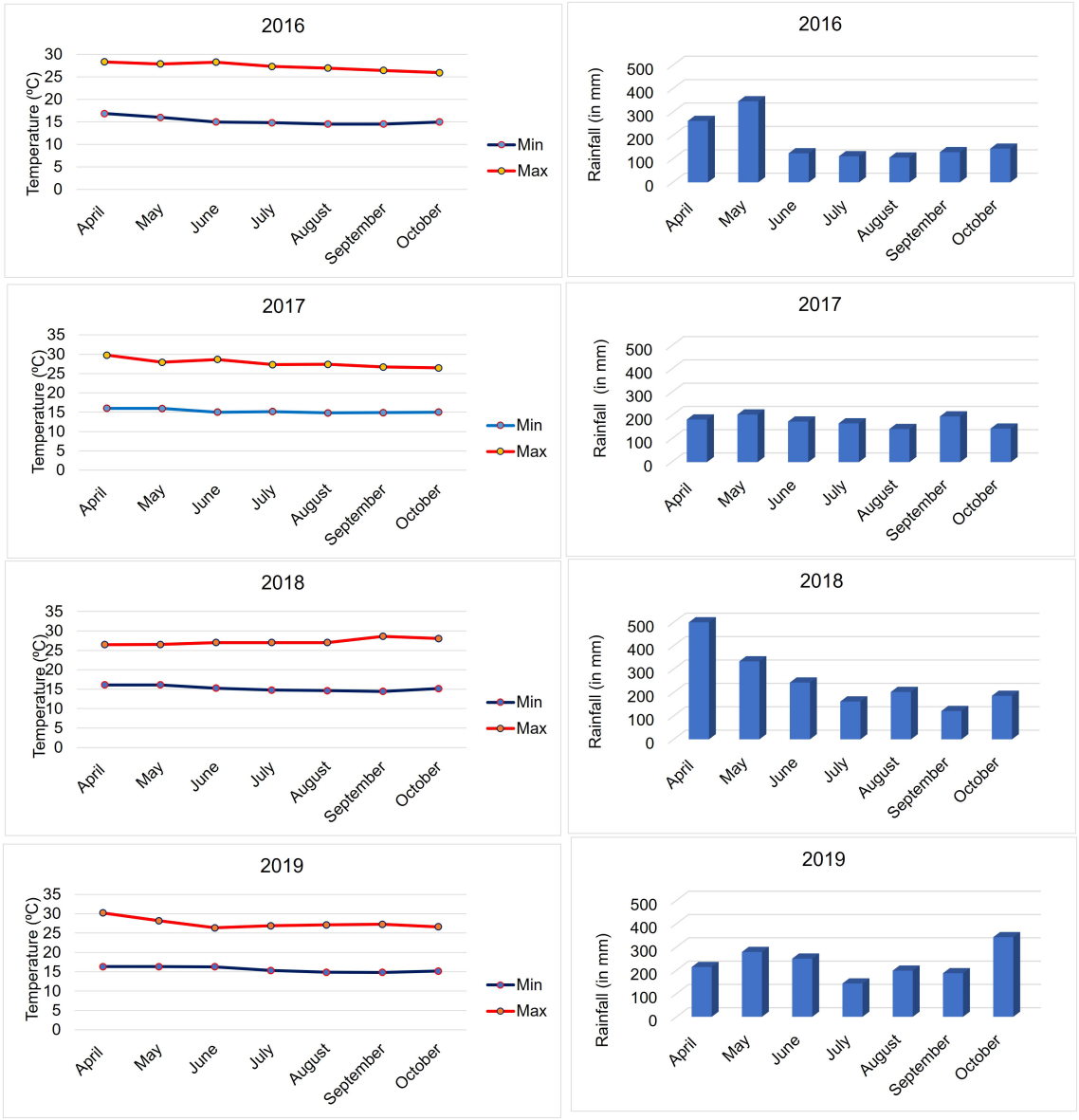


FIGURE 1
Minimum and maximum temperatures and monthly average rainfall data during the crop seasons at Kakamega for the 4 years (2016–2019) of evaluation.

where y_i = diseased leaf area estimated on the ith disease assessment date, t_i = time (days) from disease onset (i.e., inoculation) to the ith disease assessment date, and n = total number of disease assessments during the experiment evaluation. Data were also collected for other relevant agronomic traits, namely anthesis date (AD), silking date (SD), anthesis-silking interval (ASI), plant height (PH), and ear height (EH).

GLS disease severity scoring was based on an ordinal scale; therefore, data were checked for conformity with the assumptions of statistical model fitting, that is, normally distributed, constant variance, and independent (Wisniewski and Rawlings, 1990). A plot of residuals against fitted values has shown that the residuals were symmetrically distributed with constant variance for GLS disease severity data and AUDPC values; thus, the data were not transformed. Further data were assessed for homogeneity of variance using Levene's test before ANOVA, and variances were found to be homogeneous. The phenotypic traits were analyzed, and variance components were estimated with the restricted maximum likelihood (REML) in the ASREML-R (Gilmour et al., 2009) and multi-environment trial analysis (META) R software developed in CIMMYT (Alvarado et al., 2020). The following statistical model was used to estimate variance components:

$$Y_{ijkl} = \mu + G_i + E_j + (GE)_{ij} + R(\mathrm{E})_{kj} + B(RE)_{ljk} + e_{ijkl}$$

where Y_{ijkl} is the phenotypic observation at the ith inbred line, jth environment in kth replication of the lth incomplete block, μ is overall means, Gi is the genetic effect of the ith inbred line, Ej is the effect of the jth environment, $(GE)_{ij}$ is genotype by environment

interaction, $R(E)_{kj}$ is the effect of the kth replication at the jth environment, $B(RE)_{ljk}$ is the effect of the lth incomplete block in the kth replication at the jth environment, and e_{ijkl} is the residual error. META-R software (Alvarado et al., 2020) was used to obtain best linear unbiased estimates (BLUEs) and best linear unbiased predictions (BLUPs) for all traits. BLUPs were used for GWAS and BLUEs were used for GS analyses. Comparisons of variability between entries were made using the least squared differences (LSD) at a 5% significance level. Broad-sense heritability (H^2) for the different traits was calculated as the ratio of the estimated genotypic variance to the estimated phenotypic variance (Knapp et al., 1985; Liu et al., 2016).

Genotypic data analyses

From 427 lines phenotyped, we had genotypic data for 140 lines. Maize leaf tissue samples were collected from 3 to 4 weeks old young, healthy seedlings at the V3 stage. High-quality genomic DNA was isolated from freeze-dried tissues. The Diversity Array Technology (DArT) marker platform was used, and obtained 37,915 single nucleotide polymorphic (SNP) markers. TASSEL ver5.2 (Bradbury et al., 2007) was used to summarize SNP data by site, determine the allele frequencies, and implement quality screening. SNP variants that were monomorphic, called at repeat loci, had a heterozygosity of >0.05 and had a minor allele frequency of <0.05, were filtered, and 19,091 high-quality SNPs were retained for GWAS analysis.

Population structure and linkage disequilibrium

The population structure of 140 elite lines, which had both phenotypic and genotypic data, was analyzed and sub-grouped using Structure Software 2.3.4 version (Pritchard et al. 2000). The number of discontinuous population structure clusters (K) was predicted from one to five with ten iterations. The true number of population structure clusters (delta K value) was harvested online using an available structure harvester software (Earl and vonHoldt, 2012) based on the highest Ln P (D). The unique population genetic subcluster was represented by each color bar at a $p = 0.001$. The period of length of burn-in was set to 10,000, and Markov Chain Monte Carlo (MCMC) values were set to 100,000 cycles (Evanno et al., 2005).

The kinship matrix was estimated in TASSEL to measure the genetic relatedness among individuals in the association panel. The neighbor-joining tree was developed using the phylogenetic tree analysis in TASSEL software v5.2. Linkage disequilibrium (LD) was calculated using TASSEL software version 5.2. The squared allele frequency correlations (r^2) between all pairs of SNPs were estimated to determine the extent of LD. The LD decay rate was calculated by using the nonlinear regression model developed by Hill and Weir (1988), with modifications by Remington et al. (2001), was used to fit the LD decay curve into the scatterplot using the LOESS function in R.

Genome-wide association study

For GWAS, BLUPs across environments were used as phenotypes. Principal components analysis (PCA) was performed using TASSEL ver5.2 (Bradbury et al., 2007). The principal components were used to correct population structure and to create a two-dimensional plot to enable visualization of the probable population structure. A mixed linear model (MLM) that computes both PCs and a kinship matrix (K) was applied for GWAS to correct for population structure (Yu and Buckler, 2006). The R package "FarmCPU—Fixed and random model Circulating Probability Unification" was used for GWAS analysis (Liu et al., 2016). With the GAPIT package, the "hapmap" format of the markers was converted to numeric (0, 1, 2) (Wang and Zhang, 2021). The FarmCPU analysis was performed with a maxLoop of five, where the maxLoop refers to the total number of iterations used. The p threshold of 0.01 was used in the model for the first iteration, a quantitative trait nucleotide (QTN) threshold of 0.01 was used in the model from the second iteration, and a minimum MAF threshold of 0.05 was used in the analyses. To determine the significance threshold, multiple testing correction was conducted with the false discovery rate method. The significant associations were declared when p-values in independent tests were less than 3×10^{-4} (Cui et al., 2016). All the candidate genes for GLS and other agronomic traits located within regions from 5 kb upstream to 5 kb downstream associated with significant QTNs were identified and annotated using the B73 maize reference genome (B73 RefGen_V2) (Schnable et al., 2009; Slaten et al., 2020). The candidate gene annotation information was retrieved from the MaizeGDB database (http://www.maizegdb.org).

Genomic-wide prediction

Genomic prediction model, ridge-regression BLUP (RR-BLUP), was used to carry out predictions using a fivefold cross-validation (Zhao et al., 2012). BLUEs across environments were used for the analysis. A set of high-quality uniformly distributed 4,983 SNPs with no missing values and MAF > 0.05 was used. We applied a fivefold cross-validation "within population' approach, where both training and estimation sets were derived from within the association panel. The prediction accuracy was calculated as the correlation between genomic estimated breeding values (GEBVs) and the observed phenotypes. A sampling of the training and validation sets was repeated 100 times for each trait.

Results

Weather and disease incidence across the trial years

The amount of rainfall received and ambient temperatures at Kakamega varied during the crop season in all 4 years from 2016 to

2019 (Figure 1). The Kakamega region, as expected, consistently experienced high rainfall during the months of April and May, followed by a decline in precipitation during the subsequent months. However, compared to all the other three years, high rainfall and lower temperatures were observed in the months of April and May in 2018. Individual year analyses of experiments revealed significant genotypic variances (Supplementary Table S2, Supplementary Figure S1) and moderate to high repeatability for all the traits in all 4 years except for GLS1 (77 DAP) in 2017 (Supplementary Table S2). The performance of lines against GLS disease severity showed a similar distribution in the years 2016 and 2019, whereas the year 2017 was more divergent (Supplementary Figure S1). The correlation between years was positive and significant for GLS disease severity and AUDPC values (Supplementary Table S3), which supports combined analyses of the data across years.

Analysis of variance and heritability

The frequency distribution of GLS disease severity and other agronomic traits showed normal distribution (Figure 2). ANOVA across years revealed there were significant variances for genotypic and genotype x environment (GXE) interactions ($P < 0.05$) for GLS disease severity scores and agronomic traits (Table 1). The broad-sense heritability estimate for GLS1 (77 DAP) was high ($H^2 = 0.85$) compared to the GLS2 (105 DAP) score ($H^2 = 0.57$). Broad-sense heritability for agronomic traits was high for AD, SD, and PH, but moderate for EH and ASI. Three susceptible checks, namely CKL150122, CKL150079, and CKL150132 had disease severity scores of >7.3 on a 1–9 scale for GLS2. This high disease severity score on susceptible checks indicates good disease expression in the field (Table 2; Supplementary Table S4). The first GLS severity score [GLS1 (77 DAP)] varied from 1.9 to 7.0 with a mean of 4.5, while the second score [GLS2 (105 DAP)] varied from 2.5 to 7.6 with an average score of 5.7. The AUDPC ranged from 34.9 to 99.8 with a mean of 71.95. The mean performance of the lines for AD, SD, and ASI was 80, 80.4, and 0.44 days, respectively, and for PH and EH, they were 129.2 and 61.4 cm, respectively (Table 1; Figure 2).

In scoring disease severity on a 1–9 scale, scores with <4 are considered as resistant, 4–5 are moderately resistant, 5–7 are moderately susceptible, and 7–9 as completely susceptible genotypes. Among 427 lines evaluated, 125, 178, 124, and 1 line were resistant, moderately resistant, moderately susceptible, and susceptible, respectively, for GLS1 (77 DAP) disease severity score (Figure 3). For GLS2 (105 DAP), 14 lines were resistant while 69, 315, and 30 lines fell into moderately resistant, moderately susceptible, and susceptible categories (Figure 3). The 14 GLS-resistant lines also showed a wide range of diversity in their performance for agronomic traits (Table 2). The lines CML536 and CKL14501 were not only resistant to GLS but also known to be resistant to TLB (data not shown). Whereas CML566 is also known to be tolerant to drought and CML572 is tolerant to MLN. The results of the correlation analysis between the eight traits including GLS disease severity traits for the maize inbred lines are shown in Figure 4. GLS disease severity for both GLS1 (77 DAP) and GLS2

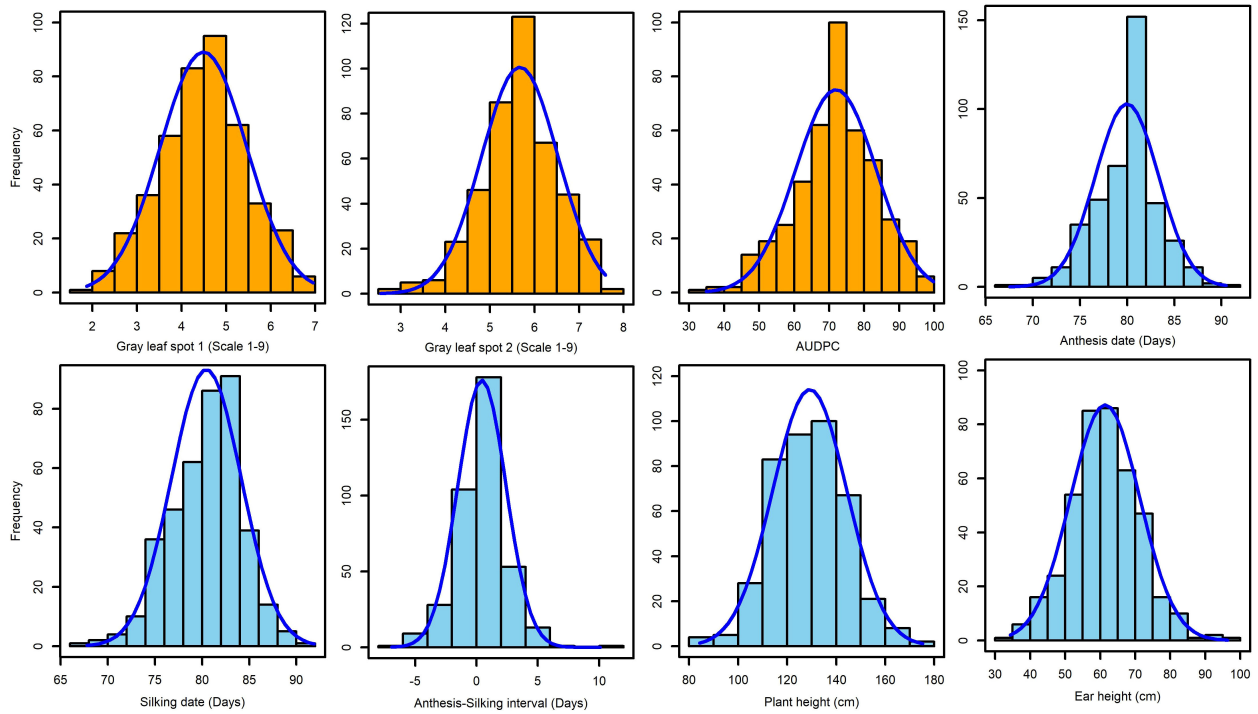


FIGURE 2
Frequency distribution of GLS disease severity and other agronomic traits. GLS1 and GLS2 = gray leaf spot disease severity data recorded at 77 and 105 days after planting, respectively.

TABLE 1 Estimation of variance components for GLS disease severity, AUDPC, and agronomic traits evaluated across 4 years (2016–2019) under artificial inoculation.

Genotype	GLS1 (1–9 scale)	GLS2 (1–9 scale)	AUDPC	AD (days)	SD (days)	ASI (days)	PH (cm)	EH (cm)
Mean	4.49	5.67	71.95	79.97	80.44	0.44	129.24	61.40
σ^{2G}	1.77**	0.40**	72.03**	8.53**	9.80**	0.88**	110.33**	40.18**
σ^{2GxE}	0.40**	0.60**	101.64**	2.26**	2.01**	1.25*	42.81**	31.14**
σ^{2e}	1.13	1.25	185.92	8.18	8.61	4.90	278.57	126.35
H^2	0.85	0.57	0.60	0.84	0.86	0.49	0.71	0.63

*, ** significance at $p = 0.05$ and 0.01 level, respectively. GLS1, GLS2, correspond to disease severity data collected at 77 and 105 days after planting, respectively; AUDPC, area under disease progress curve; AD, anthesis date; SD, silking date; ASI, anthesis silking interval; PH, plant height; EH, ear height.

(105 DAP) were significantly and negatively correlated with AD, SD, PH, and EH; however, the magnitude of correlation values was higher at the early stage of the disease severity. AUDPC values were also consistently significant and negatively correlated with AD, SD, PH, and EH. The correlation between GLS traits and ASI was non-significant. The correlation between flowering traits and PH was not significant, whereas EH was positive and significant.

Marker distribution, population structure, phylogenetic tree, and kinship

From 33,740 DART-GBS SNPs used, only 56% (19,091 SNPs) were retained after filtering with the twin criteria of >5% MAF and <10% missing per marker. The number of markers remained ranged from 1,341 on chromosome 10 to 2,876 on chromosome

TABLE 2 Disease severity scores for the best 14 lines and three susceptible checks to GLS and agronomic traits across 4 years (2016–2019).

Genotype	GLS1 (1–9 scale)	GLS2 (1–9 scale)	AUDPC	AD (days)	SD (days)	ASI (days)	PH (cm)	EH (cm)
CKDHL142989	2.5	2.6	34.9	84.0	83.4	−0.8	144.6	76.7
CKL14500	2.1	2.9	35.5	86.8	89.0	2.3	125.9	61.6
CML559	2.1	3.0	38.4	79.7	79.8	0.2	153.4	83.2
CML566	3.0	3.1	48.8	83.5	83.1	−0.2	144.6	67.7
CKL14501	3.2	3.2	45.5	85.0	85.0	−0.1	128.2	63.2
((BRAZIL1546) DH4/CML395)-B-1-2-1	2.8	3.4	49.9	90.5	91.9	1.4	126.8	71.2
CKL155	2.6	3.5	44.6	85.2	84.4	−0.7	145.3	67.8
CKDHL120423	3.0	3.6	43.2	83.0	81.8	−2.2	145.8	73.1
CKLMARS1C3S50196	3.4	3.7	49.4	79.5	81.5	2.1	125.7	61.0
CML540	3.4	3.7	49.7	75.5	75.5	0.1	137.3	49.2
CML536	3.2	3.8	59.9	85.5	85.9	1.0	136.8	61.7
CML574	2.3	3.9	45.5	83.4	82.6	−1.1	138.7	64.9
CKL14529	3.9	4.7	45.3	84.5	82.8	−1.5	142.3	68.1
CML172	4.0	4.8	45.1	74.5	78.0	3.6	118.0	53.1
CKL150079 (Suc. Check)	6.5	7.4	94.8	77.7	79.6	2.0	136.4	63.5
CKL150122 (Suc. Check)	6.2	7.4	90.0	78.2	79.9	2.8	123.7	51.3
CKL150132 (Suc. Check)	5.7	7.4	83.9	78.9	79.0	0.3	131.8	63.0
Mean	4.5	5.7	72.3	79.9	80.4	0.42	129.8	61.9
SE	0.05	0.04	0.56	0.17	0.19	0.10	0.75	0.49
$LSD_{5\%}$	1.50	1.29	17.78	4.72	5.02	2.20	20.41	13.42
CV (%)	16.92	11.49	12.52	3.00	3.18	66.17	8.00	11.04

GLS1 and GLS2 correspond to disease severity data collected for GLS at 77 and 105 days after planting, respectively; AUDPC, area under disease progress curve; AD, anthesis date; SD, silking date; ASI, anthesis silking interval; PH, plant height; EH, ear height; SE, standard error; LSD, least significant difference; CV, coefficient of variation.

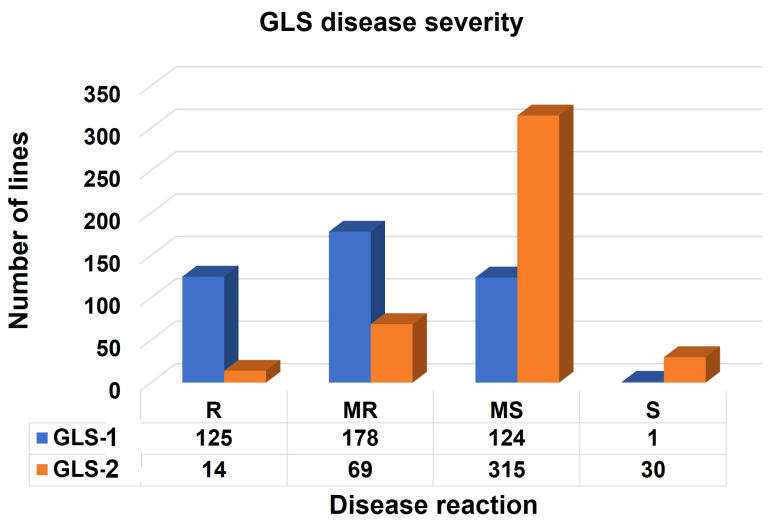


FIGURE 3
Frequency of the inbred lines with resistant (R), moderately resistant (MR), moderately susceptible (MS), and susceptible (S) reactions to GLS scouted at 77 (GLS1) and 105 (GLS2) days after planting.

1. For the final set of markers, the minimum MAF ranged between 0.05 and 0.50. The percentage of missing markers per individual varied from 0 to 10% and the overall average was 4.6%. The proportion of heterozygosity of SNPs (number of taxa that are heterozygous for a given SNP divided by the total number of individuals) ranged from 0 to 0.70, with an overall average of 0.03. The heterozygosity of inbred lines (number of heterozygous markers per inbred line divided by the total number of markers) ranged from 0.008 to 0.27 with an overall average of 0.09. The final set of 19,091 high-quality SNP markers distribution was graphically presented in Supplementary Figure S2. The relatedness among the inbred lines used for GWAS analyses was shown with the kinship matrix (Supplementary Figure S3). The population structure of 140 diverse maize lines was determined by Bayesian based model in STRUCTURE and PCA (Figure 5). The optimum number of K was obtained by plotting the number of clusters (K) against delta K which revealed delta K probability value with two and five clusters based on the highest Ln P(D) values (Figure 4A). Evanno table was constructed in the structure harvester with the highest values of 13714.1.45 Ln P(K), 127.46 standard deviations ln P(K), and 107.61 delta K. Delta K value-based line plot had suggested that the population could be structured into two and/or five groups (Figures 5B, C). The population structure was also examined by PCA which grouped all lines into two broader groups (Figure 5D). An optimal number (K) of three PCs was retained for GWAS.

The neighbor-joining method-based phylogenetic tree shown in Figure 6A revealed that the 140 diverse maize lines can be clustered into three main groups (I = 40, II = 56, and III = 44) differentiated by the different colors (Figure 6A, Supplementary Table S5). Groups I and II can also be treated as one large cluster with two sub-groups. The genome-wide LD was plotted as LD (r^2) between adjacent pairs of markers versus the distance between adjacent markers in Kb (Figure 6B). The average genome-wide LD-decay in this set of lines is 1.44 Kb at $r^2 = 0.2$. LD plots for each chromosome revealed the fastest LD-decay in chromosome 7 (0.47 Kb at $r^2 = 0.2$) and chromosome 1 displayed the slowest LD-decay (4.75 Kb at $r^2 = 0.2$).

GWAS analyses revealed 69 significant SNPs for eight traits and these results for all traits are summarized using Manhattan plots (Figure 7; Tables 3, 4) and QQ plots (Supplementary Figure S4). The Q–Q plot of the FarmCPU model showed a sharp deviation from the expected *P* value distribution in the tail area, indicating that false positives and negatives were adequately controlled (Supplementary Figure S4). Association analyses for GLS disease severity for GLS1 (77 DAP) revealed nine significant SNPs (Table 3). The allelic effect (difference in mean performance for GLS disease severity between inbred lines with major allele and minor allele) for these significant SNPs ranged from 0.45 to 0.39, −0.70 to 0.70, and −3.67 to 5.59 for GLS1 (77 DAP), GLS2 (105 DAP), and AUDPC, respectively under artificial inoculation of GLS. A negative value indicates that the minor allele was the favorable allele associated with increase in GLS resistance by decreasing the GLS severity. For GLS2 (105 DAP), that is, GLS disease severity at a late stage, four significant SNPs were detected. For AUDPC, 11 significant SNPs were identified. The significant SNPs for GLS disease severity were found on all chromosomes with the most significant one being located on chromosome 3 ($p = 1.37 \times 10^{-8}$). Information on all the significant SNPs, their corresponding MAF, and allelic effects are listed in Table 3.

For AD, a total of 13 significant SNPs were detected on chromosomes 3, 5, 6, 7, 8, and 9. One SNP on chromosome 5 (*DT5_80046482*; $p = 5.09 \times 10^{-06}$) was common between AD and GLS disease severity at a late stage. The allelic effect for these significant SNPs for AD ranged from −2.45 to 1.21. The largest number of significant SNPs as well as the most significant SNPs in the current GWAS study was identified for SD followed by AD. For SD, 14 significant SNPs distributed across all chromosomes except on chromosome 10 were identified. The SNP on chromosome 5 was

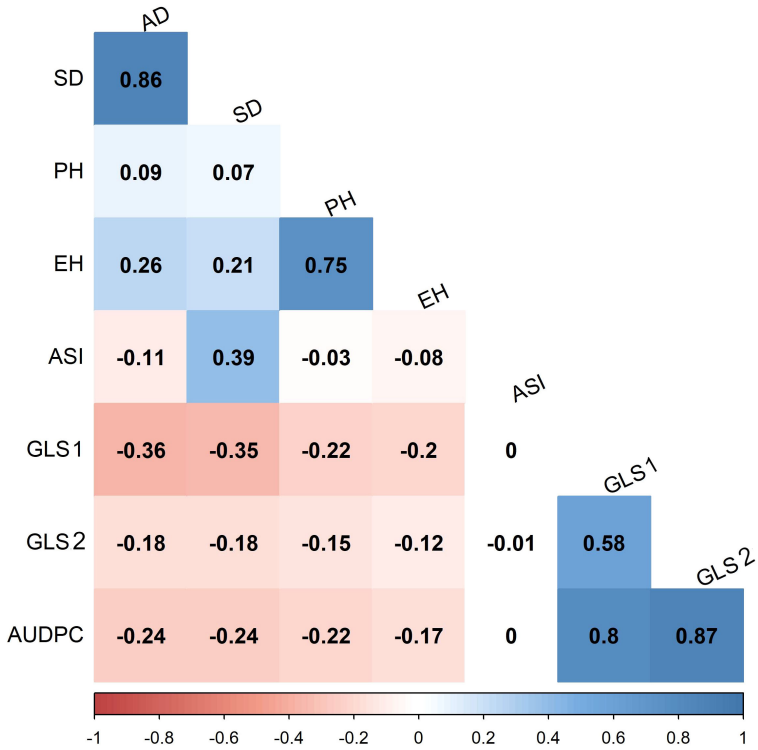


FIGURE 4
Pearson's correlation between GLS traits and other agronomic traits evaluated in four environments under artificial infestation of GLS. The correlation level is color-coded according to the color key scale plotted below. Correlations with >0.11 were significant at 0.05 (*p*) level, GLS1, GLS2, correspond to disease severity data collected for GLS at 77 and 105 days after planting, respectively; AUDPC, area under disease progress curve; AD, anthesis date; SD, silking date; ASI, anthesis silking interval; PH, plant height; EH, ear height.

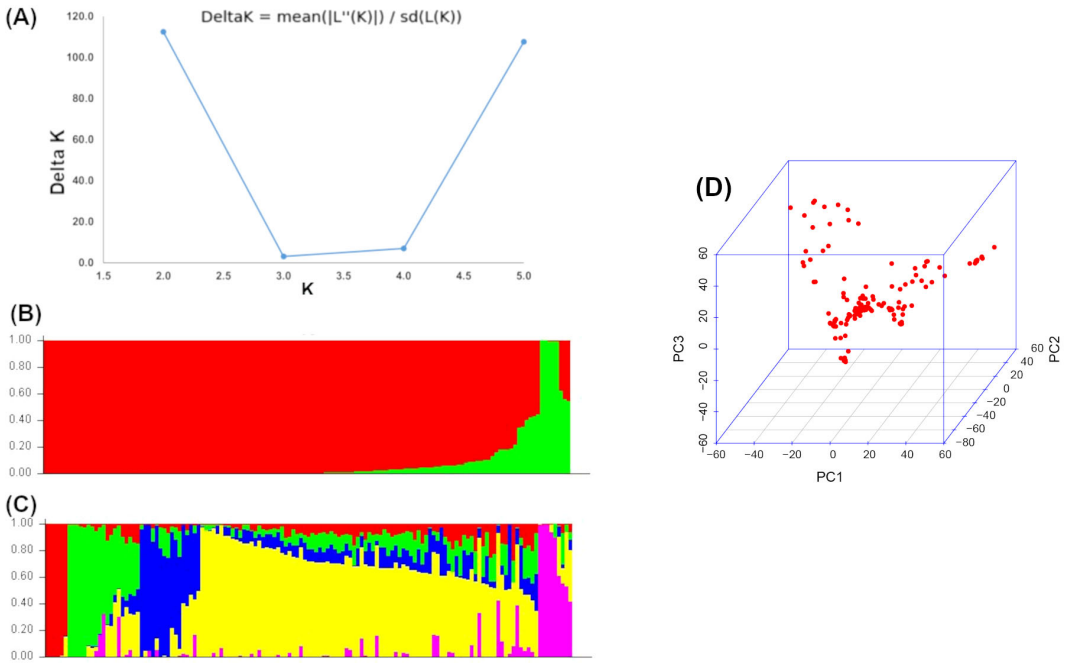


FIGURE 5
The two and five sub-populations of the 140 inbred lines by using DART SNP markers. **(A)** Best delta K estimation by Evanno method. **(B)** Estimated population structure of tropical maize inbred lines as revealed by DART SNP markers for K = 2 and **(C)** for K = 5. Blue, green, pink, red, and yellow color represents sub-population 1, 2, 3, 4, and 5, respectively. Distribution of inbred lines based on the first three principal components **(D)**.

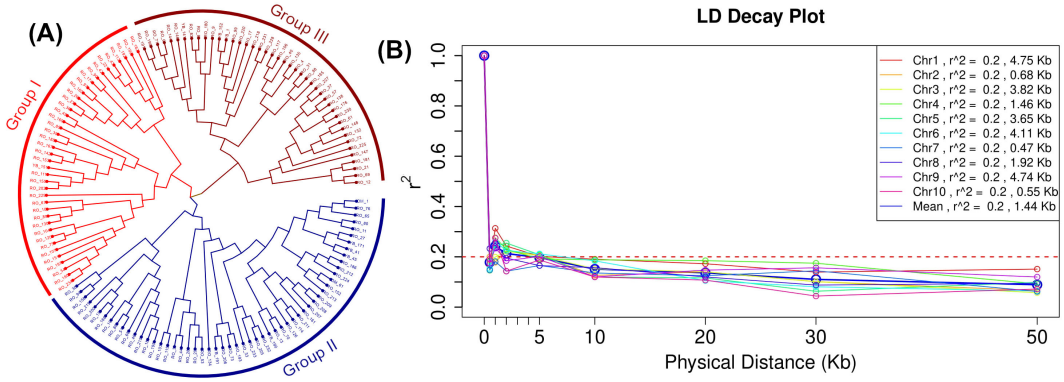


FIGURE 6
Grouping of 140 inbred lines through phylogenetic tree based on neighbor-joining method **(A)** and linkage disequilibrium (LD) plot **(B)** illustrating the average chromosome-wise and genome-wide LD decay in 140 inbred lines panel using SNPs with call rate 0.9 and minor allele frequency 0.1. The values on the Y-axis represent the squared correlation coefficient r^2 and the X-axis represents the genetic distance in kilobases (Kb).

the most significant in the current study ($p = 5.50 \times 10^{-11}$). Two SNPs on chromosome 3 (*DT3_1469626* for SD and *DT3_1407684* for AD) and another two SNPs on chromosome 9 (*DT9_76308336* for SD and *DT9_76104516* for AD) were located closely in the region (Table 4). The allelic effect for these significant SNPs for SD ranged from −1.89 to 1.65.

PH and EH are highly correlated agronomic traits in maize. There were 11 significant SNPs associated with PH distributed on chromosomes 2, 6, 8, and 10 whereas there were six significant SNPs detected for EH that were found on chromosomes 5, 7, 9, and 10. Three SNPs on chromosome 10 (*DT10_101212965*, *DT10_101212899*, and *DT10_101213049*) were common between PH and EH. The allelic effect of significant SNPs for PH ranged from −12.44 to 11.87, for EH, the range varied from −7.02 to 7.18 (Table 4).

Candidate genes analysis was conducted for significant QTNs identified in this study. A total of 20 candidate genes were discovered and annotated, among them 3 and 4 candidate genes were identified for GLS1 (77 DAP) and AUDPC values, respectively (Table 4). One candidate gene, *Zm00001d015224* is closely associated with both GLS2 (105 DAP) and AD (Table 5). Nine candidate genes potentially associated with flowering traits (AD and SD). Similarly, three and one candidate genes were found to be associated with PH and EH, respectively.

In most large advanced maize breeding programs, GS is routinely applied. Among the several GS models, G-BLUP and RR-BLUP are computationally less intensive and able to capture both major and minor effect trait variations, so they are well suited for routine application in breeding trials. Therefore, we used the RR-BLUP model to estimate the prediction accuracies in the panel for GLS and other agronomic traits. Prediction accuracies were moderate to high for all eight traits (Figure 8). The observed prediction accuracies for GLS1 (77 DAP), GLS2 (105 DAP), AUDPC value, AD, SD, ASI, PH, and EH were 0.62, 0.47, 0.50, 0.65, 0.61, 0.31, 0.34, and 0.36, respectively.

Discussion

GLS is one of the serious foliar diseases in maize (Kibe et al., 2020a; Omondi et al., 2023) and is caused by *Cercospora zeae-maydis* and *C. zeina*, in Africa *C. zeina* is more prevalent (Crous et al., 2006; Meisel et al., 2009; Liu and Xu, 2013). With the changing climate and increase in cropping the same crop over a larger area, GLS becomes a serious threat to maize production, particularly in smallholder farmers of SSA. To understand the genetics of GLS resistance, in this study, we selected a set of 427 elite lines from tropical and sub-tropical breeding pipelines adapted to eastern and southern Africa and screened them over 4 years in Kakamega under artificial inoculation of *C. zeina*.

Among the 427 lines screened for GLS, 14 were identified as resistant lines with a score of <4 on a 1–9 scale (Table 2). These lines represent both the intermediate and late maturity groups, which occupies the major market share in eastern and southern Africa. Among the 14 lines, CKDHL142989 was the best-performing line with a disease severity score of 2.6, followed by CKL14500, which showed a score of 2.9, which is promising to be used as a donor line in GLS resistance breeding. Interestingly, these lines are also tolerant to TLB, which makes them multiple disease-resistant lines. Among the selected 14 lines, CML536, CKDHL120423, CKL14500, CKL14529, and CML540 also showed resistance to TLB (data not shown), suggesting the possibility of using them as donors for resistance to both diseases. CML 536 is also known to be tolerant to drought and low soil nitrogen stress. CML 540 is resistant to MSV, TLB, common rust, early maturing, and tolerance to drought, which provide additional benefits to use these as elite lines in breeding programs. Furthermore, some of these lines may be useful as parents in abiotic stress-tolerant hybrids. For example, line CKL14500, known for GLS resistance, also carries favorable alleles for drought tolerance; it is evident as it derived from two known drought-tolerant lines (CML444 and CML395). The single cross tester (CML444 x CML395) is

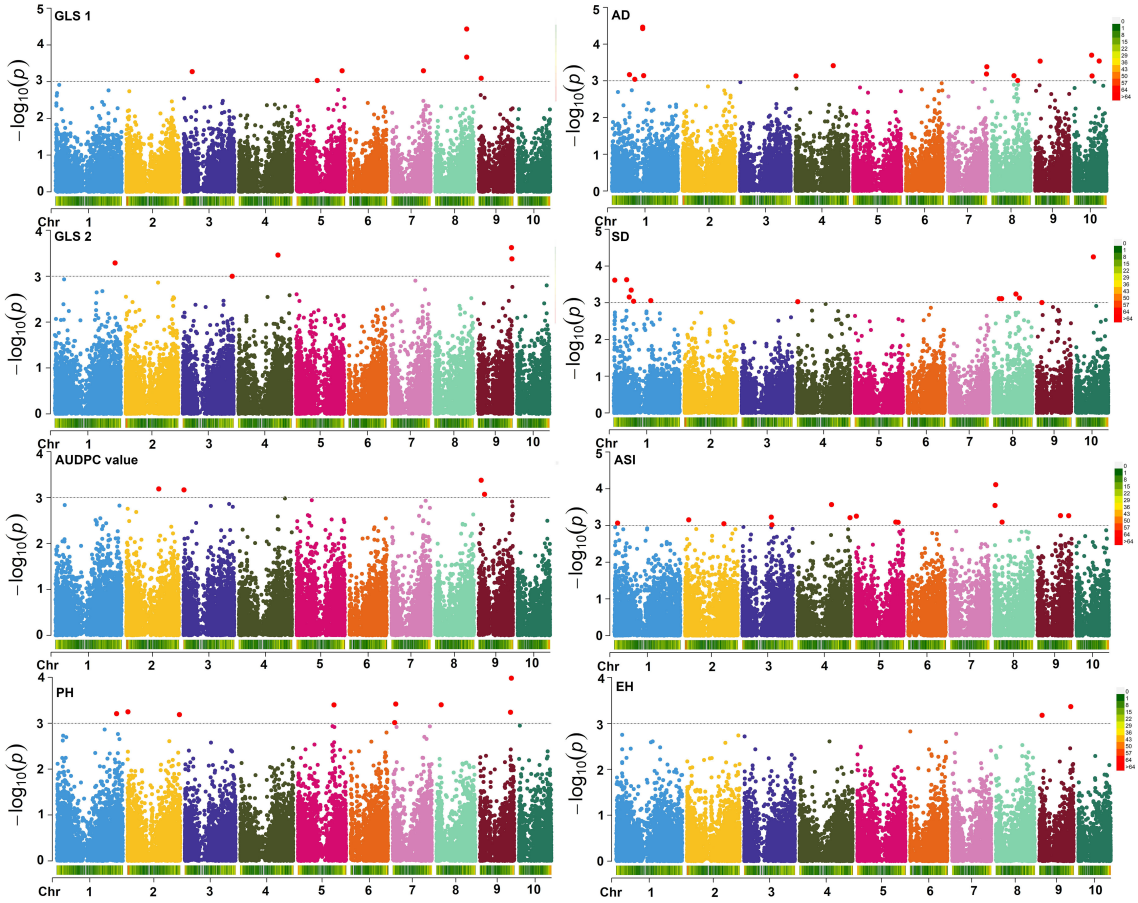


FIGURE 7
Farm CPU Manhattan plots of GWAS on GLS disease severity and other agronomic traits measured under GLS artificial inoculation. The X-axis shows the SNP position on the chromosome and Y-axis shows the negative log base 10 of the *P*-values; for ease of discrimination, each chromosome was colored differently. The horizontal line portrays the significance threshold (marker *P*-value < 3×10^{-4}). GLS1, gray leaf spot disease severity data scored 77 days after planting; GLS2, gray leaf spot disease severity data scored 105 days after planting; AUDPC, area under disease progress curve; AD, anthesis date; SD, silking date; ASI, anthesis silking interval; PH, plant height; EH, ear height.

frequently used as one of the parents for many commercially released drought-tolerant hybrids. Although line CKL14500 has some good trait combinations, it may not be a suitable parent in bi-parental crosses because of its susceptibility to common rust (Sserumaga et al., 2020). Another good line for GLS resistance (CML559) was derived from source germplasm Population 500 (P500), which is resistant to common rust, MSV, and stem borers as well as tolerant to TLB. Another line, CML566 is tolerant to drought with moderate resistance to TLB and MSV. CML574 is a yellow line and is tolerant to MLN, tar spot complex, and fall armyworm and can be used as a donor for multiple diseases and pests. Overall, the identified GLS-resistant lines not only contribute to disease resistance but also carry useful alleles for several economically important traits, including drought tolerance, which makes them multi-trait elite donors.

GLS resistance breeding is influenced by several factors, including genotype, environment, and their interactions. Across environment, analyses revealed significant genotypic and genotype x environment interaction effects for GLS disease severity and AUDPC values (Table 1). The magnitude of variance components for genotype x environment interaction variance was approximately 1.5 times of genotypic variance which indicates the role of both additive and non-additive effects. On the contrary, the magnitude of genotypic variance is more than twice that of genotype x environment interaction variance for flowering time traits like AD and SD, and for PH and EH suggesting the major role of additive effects over non-additive effects. A high magnitude of genotype x environment interaction effect was also observed in an earlier study with the IMAS association panel (Kibe et al., 2020a). The heritability estimates for these traits are also on similar expectations with moderate heritability for GLS disease traits and high heritability for flowering traits. Kibe et al. (2020a) also reported moderate heritability for GLS disease severity traits in biparental populations. GLS disease severity and AUDPC values are negatively and significantly correlated with AD, SD, PH, and EH (Figure 3). This indicates the selection of lines with early flowering and low to medium height has better resistance to GLS over lines with tall and late flowering plants. These correlations are consistent with earlier studies in association panels evaluated in SSA (Kibe et al., 2020a; Omondi et al., 2023).

10.3389/fpls.2025.1536981

TABLE 3 List of significant SNPs associated with GLS disease severity traits evaluated in four environments under artificial inoculation of *Cercospora zeina*.

SNP	Chr	Position (bp)	Bin name	MLM-P value	MAF	Effect
Gray leaf spot 1						
DT2_153752700	2	153752700	2.06	1.78×10^{-5}	0.33	−0.24
DT3_18370709	3	18370709	3.04	1.37×10^{-8}	0.29	−0.35
DT3_129473727	3	129473727	3.05	2.77×10^{-4}	0.33	0.20
DT4_236484294	4	236484294	4.09	5.49×10^{-7}	0.08	−0.45
DT5_77120507	5	77120507	5.03	4.39×10^{-6}	0.39	−0.27
DT9_45478343	9	45478343	9.03	1.67×10^{-4}	0.24	0.14
DT9_131016039	9	131016039	9.05	9.52×10^{-6}	0.50	−0.20
DT10_9375453	10	9375453	10.02	1.47×10^{-5}	0.46	0.20
DT10_124625318	10	124625318	10.04	6.96×10^{-7}	0.08	0.39
Gray leaf spot 2						
DT3_8260307	3	8260307	3.03	8.30×10^{-5}	0.10	0.70
DT5_80046482	5	80046482	5.03	2.04×10^{-4}	0.30	−0.70
DT6_118742764	6	118742764	6.04	1.50×10^{-4}	0.32	0.40
DT10_7310310	10	7310310	10.02	2.13×10^{-4}	0.24	−0.45
AUDPC value						
DT1_209722581	1	209722581	1.07	2.18×10^{-4}	0.36	−2.32
DT1_257413254	1	257413254	1.09	4.11×10^{-5}	0.12	5.01
DT2_55324276	2	55324276	2.04	1.41×10^{-4}	0.31	−2.93
DT3_226339092	3	226339092	3.09	8.44×10^{-7}	0.44	−3.67
DT5_24851058	5	24851058	5.03	1.43×10^{-5}	0.15	5.59
DT6_1389517	6	1389517	6.00	1.47×10^{-5}	0.22	−3.46
DT6_91304345	6	91304345	6.02	1.32×10^{-4}	0.25	−2.81
DT6_165363652	6	165363652	6.08	2.06×10^{-4}	0.12	3.44
DT7_100370211	7	100370211	7.02	4.39×10^{-5}	0.42	−2.57
DT8_15011837	8	15011837	8.02	9.07×10^{-6}	0.23	3.83
DT9_138867161	9	138867161	9.05	1.24×10^{-5}	0.50	−2.83

MAF, Minor Allele Frequency, effect- Allele Effect, MLM P-value- probability value for the mixed linear model, GLS disease severity, and AUDPC- area under disease progress curve values under artificial inoculation of GLS conditions; Position - The physical position of the SNP (Ref Gen_v3 of B73).

GWAS results are influenced by population structure, as it influences marker-trait associations, including the false positives in an association mapping panel. We observed low to moderate population structure with PC1 and PC2 explaining 6.53% and 5.56% of variation, respectively. The association panel lines are broadly grouped into two clusters. This was also supported by phylogenetic tree grouping, though three groups were formed, a closer look shows groups 1 and 2 form one major cluster (Figure 6A). Similar findings were also observed by Kibe et al. (2020a) in an association panel and Zhao et al. (2012) in a large set of DH lines which comprised tropical and subtropical lines. Further to understand the structure of the panel used in this study, STRUCTURE software was used, where the *ad hoc* statistics ΔK were used to determine the optimum number of subgroups based on the output log likelihood of data [LnP (D)]. The peaks of the line plot (Figure 5A) suggest that the GWAS panel could be broadly divided into two groups or five subgroups. The kinship matrix also suggests moderate structure among the lines used for GWAS (Supplementary Figure S3). The moderate structure observed in the panel with no clear differentiation of major adaptation groups is in anticipation of earlier studies where CIMMYT maize germplasm was not found to have a strong population structure (Wu et al., 2016; Rashid et al., 2020). Several researchers have also reported moderate structure in the tropical maize germplasm (Kibe et al., 2020a; Nyaga et al., 2020; Sadessa et al., 2022). CIMMYT's germplasm pools and populations are known for high genetic

10.3389/fpls.2025.1536981

TABLE 4 List of significant SNPs associated with agronomic traits evaluated in four environments under artificial inoculation of *Cercospora zeina*.

SNP	Chr	Position (bp)	Bin name	MLM-P value	MAF	Effect
Anthesis date						
DT3_1407684	3	1407684	3.01	2.03×10^{-6}	0.34	−0.72
DT3_28537024	3	28537024	3.04	1.25×10^{-4}	0.05	−1.51
DT3_159134665	3	159134665	3.05	1.33×10^{-4}	0.06	−1.25
DT3_173798390	3	173798390	3.06	2.39×10^{-5}	0.35	−0.74
DT5_4115653	5	4115653	5.01	2.18×10^{-6}	0.15	1.04
DT5_80046482	5	80046482	5.03	5.09×10^{-6}	0.30	1.21
DT5_190321766	5	190321766	5.05	1.18×10^{-5}	0.35	−0.70
DT6_114255122	6	114255122	6.04	7.73×10^{-9}	0.21	1.18
DT6_146530203	6	146530203	6.05	1.68×10^{-5}	0.30	0.71
DT7_168724486	7	168724486	7.05	2.92×10^{-8}	0.35	−1.00
DT8_148947397	8	148947397	8.06	1.14×10^{-8}	0.11	−1.43
DT9_76104516	9	76104516	9.03	1.16×10^{-10}	0.06	−2.45
DT9_116598435	9	116598435	9.04	1.88×10^{-4}	0.42	0.64
Silking date						
DT1_36755457	1	36755457	1.03	9.22×10^{-5}	0.16	−0.76
DT1_160031116	1	160031116	1.05	3.41×10^{-5}	0.29	−0.76
DT1_196931772	1	196931772	1.06	2.47×10^{-5}	0.22	−0.84
DT2_215851764	2	215851764	2.08	2.22×10^{-6}	0.07	−1.31
DT3_1469626	3	1469626	3.01	6.50×10^{-5}	0.50	0.61
DT3_9577871	3	9577871	3.03	8.45×10^{-7}	0.44	−0.82
DT3_132334454	3	132334454	3.05	3.16×10^{-6}	0.07	−1.38
DT4_208901175	4	208901175	4.09	1.29×10^{-4}	0.06	−1.22
DT5_217466828	5	217466828	5.09	5.50×10^{-11}	0.16	−1.47
DT6_85907826	6	85907826	6.01	6.26×10^{-5}	0.06	1.65
DT7_54538519	7	54538519	7.02	2.60×10^{-5}	0.49	−0.61
DT7_174089319	7	174089319	7.05	1.95×10^{-4}	0.13	−0.78
DT8_63984569	8	63984569	8.03	3.34×10^{-5}	0.15	−0.93
DT9_76308336	9	76308336	9.03	4.59×10^{-6}	0.08	−1.89
Anthesis-Silking interval						
DT5_216421061	5	216421061	5.09	2.09×10^{-4}	0.21	0.71
DT7_7541489	7	7541489	7.01	3.02×10^{-4}	0.05	1.18
Plant height						
DT2_11667280	2	11667280	2.02	1.28×10^{-4}	0.14	7.93
DT2_27642609	2	27642609	2.03	1.45×10^{-4}	0.39	5.16
DT2_27774525	2	27774525	2.03	1.40×10^{-4}	0.41	5.11
DT6_168920192	6	168920192	6.08	3.31×10^{-5}	0.06	−12.44
DT8_122249319	8	122249319	8.04	1.69×10^{-4}	0.09	−8.72

(Continued)

TABLE 4 Continued

SNP	Chr	Position (bp)	Bin name	MLM-P value	MAF	Effect
Plant height						
DT8_138595600	8	138595600	8.05	2.14×10^{-4}	0.18	−6.60
DT10_56267943	10	56267943	10.03	6.73×10^{-5}	0.05	11.87
DT10_101212899	10	101212899	10.04	8.59×10^{-5}	0.05	−11.54
DT10_101212965	10	101212965	10.04	7.99×10^{-5}	0.06	10.78
DT10_101213049	10	101213049	10.04	1.97×10^{-5}	0.07	10.99
DT10_124620967	10	124620967	10.04	1.99×10^{-4}	0.10	−8.19
Ear height						
DT5_215513759	5	215513759	5.08	3.32×10^{-5}	0.12	−5.54
DT7_5660772	7	5660772	7.01	1.61×10^{-4}	0.14	4.76
DT9_127735255	9	127735255	9.05	6.61×10^{-5}	0.21	4.21
DT10_101212899	10	101212899	10.04	2.20×10^{-4}	0.05	−7.02
DT10_101212965	10	101212965	10.04	4.18×10^{-5}	0.06	7.18
DT10_101213049	10	101213049	10.04	7.69×10^{-5}	0.07	6.60

MAF, minor allele frequency, effect = allele effect; MLM *P*-value = probability value for the mixed linear model; GLS, disease severity, and AUDPC, area under disease progress curve values under artificial inoculation of GLS conditions; position = the physical position of the SNP (Ref Gen_v3 of B73).

diversity and serve as the source of germplasm for many breeding lines in the tropical and subtropical regions (George et al., 2004; Warburton et al., 2005). This is one of the possible reasons we did not observe any well-defined population structure in this study.

The rate of LD decay indicates the presence of diversity in the selected germplasm or panels. Fast LD decay suggests higher diversity at the nucleotide base level, which might have resulted from the historic recombination events. In temperate maize germplasm, LD decay distance (10–100 kb) is several times higher than that of tropical maize germplasm (5–10 kb, Lu et al., 2011). Romay et al. (2013) found that LD decays much more rapidly in the tropical germplasm to about 1 kb at $r^2 = 0.2$. A rapid LD decay was observed in each chromosome and across the genome in the panel (1.44 Kb at $r^2 = 0.2$). The LD decay observed in this study corroborates earlier studies (Rashid et al., 2020; Kibe et al., 2020a), which suggests the presence of sufficient diversity in the selected set of lines for GWAS.

GWAS revealed 13 markers significantly associated with GLS disease severity and 11 markers with AUDPC value (Table 3). Interestingly, though GLS disease severity was positively and significantly correlated with AUDPC values, no common SNPs were detected across GLS traits in GWAS analyses. Nevertheless, there are few regions overlapped in terms of their bin locations (Table 3). Many studies reported QTLs for GLS resistance are distributed in all 10 chromosomes, where most of them explained small to moderate effects, except for a very few that had a major effect of >10% and were used for further fine mapping studies (Du et al., 2020; Zhu et al., 2021; Kibe et al., 2020a; Omondi et al., 2023). For three foliar diseases including, GLS, TLB, and SLB revealed 147 multiple disease resistance mQTLs through meta-QTL analyses and identified bins 3.04–08, 5.04–07, and 8.05–06 as significant regions for resistance to these diseases (Ali et al., 2013). Summarizing the earlier QTL studies revealed five major clusters or hot spots for GLS resistance, namely in chromosome 1 at bin's 1.05–1.06 (Saghai Maroof et al., 1996; Lehmensiek et al., 2001; Balint-Kurti et al., 2008; Pozar et al., 2009; Xu et al., 2014), on chromosome 2 at bin's 2.03–2.05 (Bubeck et al., 1993; Saghai Maroof et al., 1996; Zwonitzer et al., 2010; Zhang et al., 2012; Lennon et al., 2016), in chromosome 4 at bin's 4.05–4.08 (Bubeck et al., 1993; Saghai Maroof et al., 1996; Clements et al., 2000; Balint-Kurti et al., 2008; Zwonitzer et al., 2010; Zhang et al., 2012; Benson et al., 2015; Lennon et al., 2016), in chromosome 5 at bin's 5.03–5.06 (Bubeck et al., 1993; Clements et al., 2000; Lehmensiek et al., 2001; Zhang et al., 2012; Lennon et al., 2016) and in chromosome 7 at bin's 7.02–7.03 (Bubeck et al., 1993; Pozar et al., 2009; Zwonitzer et al., 2010; Berger et al., 2014; Benson et al., 2015; Mammadov et al., 2015). Comparison of these hot spots with our results revealed two markers on chromosome 2 (*DT2_55324276; DT2_153752700*), three markers on chromosome 5 (*DT5_77120507; DT5_80046482; DT5_24851058*), one marker on chromosome 7 (*DT7_100370211*) were co-located within these regions. This supports the earlier findings and indicates their consistent association with GLS resistance in different genetic backgrounds. These regions are of potential interest to identify possible potential candidate genes and use them to improve GLS resistance.

The other SNPs, though not fallen into these hotspot regions, are overlapped with a few reported QTLs from other studies (Du et al., 2020; Kibe et al., 2020a; Chen et al., 2021; Qiu et al., 2021; Zhu et al., 2021). For instance, the most significant SNP associated with GLS disease severity in this study was *DT3_18370709* at the physical position 18.37 Mb on chromosome 3 ($P = 1.37 \times 10^{-8}$) was

overlapped with "consensus QTL" on bin 3.04 in the IBM2005 map (Shi et al., 2007), and the QTL (*qYCM-DS3-1*) reported in the RIL population (Chen et al., 2021). Another SNP (*DT3_129473727*) at bin 3.04 is overlapped with an earlier reported SNP through GWAS (Kuki et al., 2018) and QTL mapping (Du et al., 2020). All three SNPs detected in this study were also overlapped with an earlier study by (Kibe et al., 2020) in the GWAS panel and biparental populations. The remaining markers, especially on chromosomes 9 and 10, appeared to be specific for the current study and new additional sources for GLS resistance.

Flowering traits, both AD and SD, also play crucial roles in selecting GLS resistance. SNP (*DT5_80046482*) at chromosome 5 is significantly associated with AD and also showed a strong association with GLS disease severity (Tables 3, 4). Further, we also observed eight bins, namely, bins 3.03, 3.04, 3.05, 4.09, 5.03, 6.04, 7.02, and 9.03 shared SNPs for both GLS disease severity and flowering time traits (Tables 3, 4). Other agronomic traits like PH and EH are also significantly and negatively correlated with GLS resistance. A comparison of markers detected for GLS disease severity and PH and EH revealed three bins, bins 6.08, 9.05, and 10.04 shared the markers. These findings suggest strong linkage or clustering of markers in certain regions of the genome and selecting early flowering time, and medium height also indirectly helps in improving GLS resistance.

Among the eight candidate genes identified in this study for GLS resistance, one on chromosome 5 (*Zm00001d015224*) associate with both GLS and AD encodes for Salicylate/benzoate carboxyl methyltransferase (Table 5). Salicylate carboxyl methyltransferase responsible for formation of methyl salicylate which plays important role in signaling for local defense and systemic acquired resistance of plants against pathogens invasion (Koo et al., 2007). Another candidate gene associated with AUDPC value on chromosome 5 (*Zm00001d013920*) encodes for Phosphatidylinositol 4-phosphate 5-kinase which plays a role in plant defense and cellular function (Zarreen et al., 2023). Another candidate gene encodes for receptor homology region transmembrane domain- and RING domain-containing protein involved in transport of storage proteins to protein storage vacuoles. Overall, the identified candidate genes in this study are involved in plant defense and development.

TABLE 5 Candidate genes for GLS disease severity traits and other agronomic traits under artificial inoculation of GLS conditions.

Trait	SNP	Chr	Position (bp)	Gene_name	Annotation
GLS1 (77 DAP)	DT4_236484294	4	236484294	Zm00001d053613	RNA-binding (RRM/RBD/RNP motifs) family protein
GLS1 (77 DAP)	DT9_45478343	9	45478343	Zm00001d045883	ADP/ATP carrier protein 1 mitochondrial
GLS1 (77 DAP)	DT10_9375453	10	9375453	Zm00001d023539	protein_coding
GLS2 (105 DAP)	DT5_80046482	5	80046482	Zm00001d015224	Salicylate/benzoate carboxyl methyltransferase
AUDPC	DT5_24851058	5	24851058	Zm00001d013920	Phosphatidylinositol 4-phosphate 5-kinase 9
AUDPC	DT6_1389517	6	1389517	Zm00001d034998	Receptor homology region transmembrane domain- and RING domain-containing protein 2
AUDPC	DT8_15011837	8	15011837	Zm00001d008623	Very-long-chain (3R)-3-hydroxyacyl-CoA dehydratase PASTICCINO 2
AUDPC	DT9_138867161	9	138867161	Zm00001d047673	DUF4378 domain protein
AD	DT3_1407684	3	1407684	Zm00001d039305	RNA polymerase I-associated factor PAF67
AD	DT5_80046482	5	80046482	Zm00001d015224	Salicylate/benzoate carboxyl methyltransferase
AD	DT6_114255122	6	114255122	Zm00001d037159	Os05g0597150 protein
AD	DT9_76104516	9	76104516	Zm00001d046243	Putative ferroportin-domain family protein
SD	DT1_36755457	1	36755457	Zm00001d028486	Chemocyanin
SD	DT1_160031116	1	160031116	Zm00001d030795	Plant calmodulin-binding protein-related
SD	DT3_1469626	3	1469626	Zm00001d039313	protein_coding
SD	DT6_85907826	6	85907826	Zm00001d036370	protein_coding
SD	DT8_63984569	8	63984569	Zm00001d009426	Mov34/MPN/PAD-1 family protein
PH	DT2_11667280	2	11667280	Zm00001d002390	DNA ligase 4
PH	DT2_27642609	2	27642609	Zm00001d002942	Tubulin-folding cofactor B
PH	DT2_27774525	2	27774525	Zm00001d002944	Dolichyl-diphosphooligosaccharide–protein glycosyltransferase 67 kDasubunit
EH	DT7_5660772	7	5660772	Zm00001d018802	protein_coding

GLS1 and GLS2 correspond to disease severity data collected for GLS at 77 and 105 days after planting, respectively; AUDPC, area under disease progress curve; AD, anthesis date; SD, silking date; PH, plant height; EH, ear height.

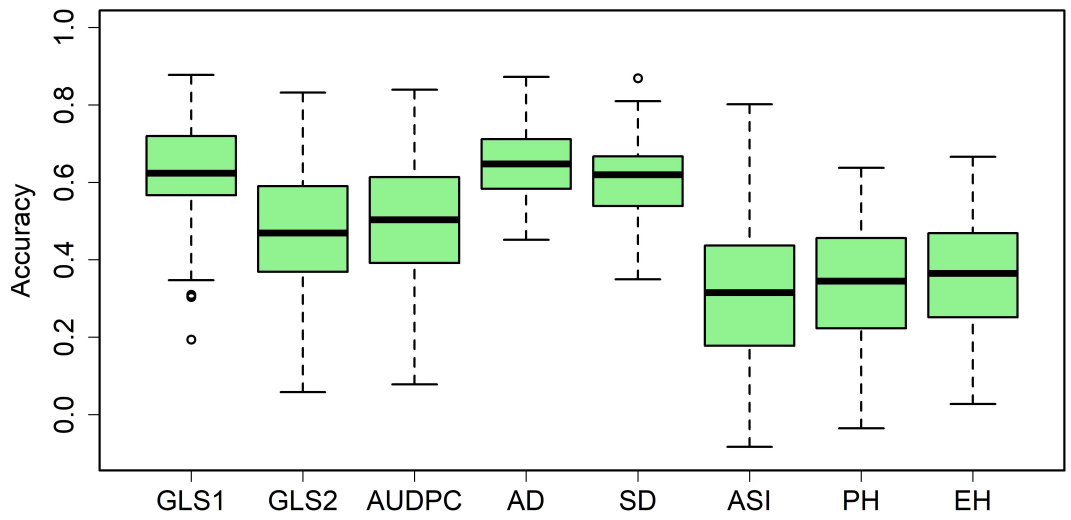


FIGURE 8
Box-whisker plots for the accuracy of genomic predictions assessed by fivefold cross-validation within the association mapping panel. GLS1 = gray leaf spot disease severity data scored 77 days after planting; GLS2 = gray leaf spot disease severity data scored 105 days after planting; AUDPC, area under the disease progress curve; AD, days to anthesis; SD, days to silking; PH, plant height; EH, ear height.

GLS resistance-linked QTLs are reported across all chromosomes. Some of the major effects of QTLs on chromosomes 1, 2, 5, and 8 were also fine-mapped (Liu et al., 2016; Zhang et al., 2017; Du et al., 2020; Qiu et al., 2021). As the number of QTL needed to be considered to achieve maximum resistance to GLS increases, it complicates the success of marker-assisted selection. Unlike GWAS, which identifies the markers linked to the trait of interest, GS calculates the estimated breeding values of the trait/s for genotypes in practical breeding which is used to select superior-performing candidates. GS has a high predictive power since it uses genome-wide markers to predict the breeding value of individuals in the testing population (Meuwissen et al., 2001). The rapid innovation in next-generation sequencing technology able to produce millions of markers with reduced cost of genotyping makes GS a critical method in breeding programs. Currently in maize, GS is used for complex traits like grain yield and drought tolerance (Zhao et al., 2012; Beyene et al., 2019, 2021b). Compared to abiotic stress traits and grain yield, diseases like GLS are relatively less complex, following this expectation, we observed a high prediction correlation for GLS disease severity and a moderate correlation for AUDPC values (Figure 8). These correlations are higher than observed correlations in IMAS association panel for GLS disease severity but on par with correlations observed in DH populations (Kibe et al., 2020a; Omondi et al., 2023). GS on several disease traits showed promising results with a prediction accuracy of as high as 0.70 for TLB (Technow et al., 2013) 0.86 for MLN (Sitonik et al., 2019), and moderate accuracy of 0.46 for Fusarium ear rot resistance (Holland et al., 2020; Kuki et al., 2020) and common rust resistance (Kibe et al., 2020b; Nyaga, et al., 2020). By using a large, related, and improved training population, the prediction accuracy could be greatly elevated as shown for GLS in the previous study, GS accuracy was low-to-moderate with a range of 0.29–0.56 for GLS resistance with a small training population, which was elevated to 0.77 when increasing the diversity and size of the training set (Kibe et al., 2020). Therefore, complementing GS with phenotypic selection is promising to achieve high genetic gain for GLS resistance with optimal resources.

Conclusion

In this study, 427 diverse tropical maize inbred lines were evaluated in 4 years under artificial inoculation of GLS (*Cercospora zeina*) in Kakamega, Kenya. Wide variation was observed for GLS disease severity and AUDPC values across years, but significantly and negatively correlated with agronomic traits such as flowering time and plant height. We identified fourteen GLS-resistant lines that can be used as either donors or parents in a resistance hybrid breeding program. In SSA, three-way cross hybrids are the final commercial products. Therefore, using lines with moderate resistance to GLS can combine the desirable alleles from all three parents, which also adds up to higher resistant hybrids. From 427 lines a set of 140 lines were genotyped with DART GBS genotyping. Population structure analyses revealed moderate structure in the panel. GWAS analyses revealed 24 SNPs significantly associated with GLS traits. Most of the detected SNPs were also co-located with earlier studies for GLS resistance. This indicates the consistency in the detection of the markers across genetic backgrounds. GS prediction correlations are moderate to high, which opens new avenues to improve breeding for GLS disease resistance with optimum allocation of resources.

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

LS: Conceptualization, Data curation, Formal analysis, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. MG: Formal analysis, Investigation, Project administration, Software, Supervision, Visualization, Writing – original draft, Writing – review & editing. YB: Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing. DM: Data curation, Investigation, Methodology, Validation, Visualization, Writing – review & editing. MK: Formal analysis, Software, Validation, Writing – review & editing. JB: Formal analysis, Methodology, Software, Validation, Visualization, Writing – review & editing. RO: Data curation, Investigation, Validation, Visualization, Writing – review & editing. VW: Data curation, Investigation, Methodology, Visualization, Writing – review & editing. BP: Conceptualization, Funding acquisition, Project administration, Resources, Supervision, 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 supported by the Bill and Melinda Gates Foundation (BMGF), Foundation for Food and Agriculture Research (FFAR), and the United States Agency for International Development (USAID) through AG2MW (Accelerating Genetic Gains in Maize and Wheat for Improved Livelihoods, BMGF Investment ID INV-003439), and the CGIAR Research Program on Maize (MAIZE).

Acknowledgments

The authors are grateful to the International Maize and Wheat Improvement Center (CIMMYT) and partner scientists and technicians who generated the germplasm, designed and conducted the experiments at KALRO research station, Kakamega, Kenya which we used to explore our objectives. The authors highly appreciate the technical support from the staff members affiliated with CIMMYT maize research station in Kiboko, Kenya.

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

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.2025.1536981/full#supplementary-material

References

Ali, F., Ali, F., Pan, Q., Chen, G., Zahid, K. R., and Yan, J. (2013). Evidence of multiple disease resistance (MDR) and implication of meta-analysis in marker assisted selection. *PloS One* 8 (7), e68150. doi: 10.1371/journal.pone.0068150

Alvarado, G., Rodríguez, F. M., Pacheco, A., Burgueño, J., Crossa, J., Vargas, M., et al. (2020). META-R: A software to analyze data from multi-environment plant breeding trials. *Crop J.* 8, 745–756. doi: 10.1016/j.cj.2020.03.010

Balint-Kurti, P. J., Zwonitzer, J. C., Pè, M. E., Pea, G., Lee, M., and Cardinal, A. J. (2008). Identification of quantitative trait loci for resistance to southern leaf blight and days to anthesis in two maize recombinant inbred line populations. *Phytopathology®* 98, 315–320. doi: 10.1094/PHYTO-98-3-0315

Benson, J. M., Poland, J. A., Benson, B. M., Stromberg, E. L., and Nelson, R. J. (2015). Resistance to gray leaf spot of maize: genetic architecture and mechanisms elucidated through nested association mapping and near-isogenic line analysis. *PloS Genet.* 11, e1005045. doi: 10.1371/journal.pgen.1005045

Berger, D. K., Carstens, M., Korsman, J. N., Middleton, F., Kloppers, F. J., Tongoona, P., et al. (2014). Mapping QTL conferring resistance in maize to gray leaf spot disease caused by Cercospora zeina. *BMC Genet.* 15, 60. doi: 10.1186/1471-2156-15-60

Beyene, Y., Gowda, M., Olsen, M., Robbins, K. R., Pérez-Rodríguez, P., Alvarado, G., et al. (2019). Empirical comparison of tropical maize hybrids selected through genomic and phenotypic selections. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.01502

Beyene, Y., Gowda, M., Pérez-Rodríguez, P., Olsen, M., Robbins, K. R., Burgueño, J., et al. (2021). Application of genomic selection at the early stage of breeding pipeline in tropical maize. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.685488

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

Bubeck, D. M., Goodman, M. M., Beavis, W. D., and Grant, D. (1993). Quantitative trait loci controlling resistance to gray leaf spot in maize. *Crop Sci.* 33, 838–847. doi: 10.2135/cropsci1993.0011183X003300040041x

Chen, L., Liu, L., Li, Z., Zhang, Y., Kang, M. S., Wang, Y., et al. (2021). High-density mapping for gray leaf spot resistance using two related tropical maize recombinant inbred line populations. *Mol. Biol. Rep.* 48, 3379–3392. doi: 10.1007/s11033-021-06350-9

Clements, M. J., Dudley, J. W., and White, D. G. (2000). Quantitative trait loci associated with resistance to gray leaf spot of corn. *Phytopathology* 90, 1018–1025. doi: 10.1094/PHYTO.2000.90.9.1018

Crous, P. W., Groenewald, J. Z., Groenewald, M., Caldwell, P., Braun, U., and Harrington, T. C. (2006). Species of Cercospora associated with grey leaf spot of maize. *Stud. Mycology* 55, 189–197. doi: 10.3114/sim.55.1.189

Cui, Z., Luo, J., Qi, C., Ruan, Y., Li, J., Zhang, A., et al. (2016). Genome-wide association study (GWAS) reveals the genetic architecture of four husk traits in maize. *BMC Genomics* 17, 946. doi: 10.1186/s12864-016-3229-6

Dodd, J., and McGee, D. C. (1989). Maize diseases: A reference source for seed technologists. *Mycologia* 81 (4), 493–494. doi: 10.2307/3760093

Du, L., Yu, F., Zhang, H., Wang, B., Ma, K., Yu, C., et al. (2020). Genetic mapping of quantitative trait loci and a major locus for resistance to grey leaf spot in maize. *Theor. Appl. Genet.* 133, 2521–2533. doi: 10.1007/s00122-020-03614-z

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

Erenstein, O., Jaleta, M., Sonder, K., Mottaleb, K., and Prasanna, B. M. (2022). Global maize production, consumption and trade: trends and R&D implications. *Food Sec.* 14, 1295–1319. doi: 10.1007/s12571-022-01288-7

Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Mol. Ecol.* 14, 2611–2620. doi: 10.1111/j.1365-294X.2005.02553.x

FAOSTAT (2021). *Food and Agriculture Organization Corporate Statistical Database*. Available online at: http://www.fao.org/faostat/en/data/QC (Accessed September 15, 2024).

George, M. L. C., Regalado, E., Li, W., Cao, M., Dahlan, M., Pabendon, M., et al. (2004). Molecular characterization of Asian maize inbred lines by multiple laboratories. *Theor. Appl. Genet.* 109, 80–91. doi: 10.1007/s00122-004-1626-8

Gilmour, A. R., Gogel, B. J., Cullis, B. R., Thompson, R., Butler, D., Cherry, M., et al. (2009). "ASReml user guide release 3.0," (Hemel Hempstead, UK: VSN International Ltd).

Gowda, M., Das, B., Makumbi, D., Babu, R., Semagn, K., Mahuku, G., et al. (2015). Genome-wide association and genomic prediction of resistance to maize lethal necrosis disease in tropical maize germplasm. *Theor. Appl. Genet.* 128, 1957–1968. doi: 10.1007/s00122-015-2559-0

Guo, T., Yu, X., Li, X., Zhang, H., Zhu, C., Flint-Garcia, S., et al. (2019). Optimal designs for genomic selection in hybrid crops. *Mol. Plant* 12, 390–401. doi: 10.1016/j.molp.2018.12.022

Hill, W. G., and Weir, B. S. (1988). Variances and covariances of squared linkage disequilibria in finite populations. *Theor. Popul. Biol.* 33, 54–78. doi: 10.1016/0040-5809(88)90004-4

Holland, J. B., Marino, T. P., Manching, H. C., and Wisser, R. J. (2020). Genomic prediction for resistance to Fusarium ear rot and fumonisin contamination in maize. *Crop Sci.* 60, 1863–1875. doi: 10.1002/csc2.20163

Katwal, T. B., Wangchuk, D., Dorji, L., Wangdi, N., and Choney, R. (2013). Evaluation of gray leaf spot tolerant genotypes from CIMMYT in the highland maize production eco-systems of Bhutan. *J. Life Sci.* 7, 443–452.

Kerns, M. R., Dudley, J. W., and Rufener, G. K. (1999). QTL for resistance to common rust and smut in maize'. *Maydica* 44, 1.

Kibe, M., Nair, S. K., Das, B., Bright, J. M., Makumbi, D., Kinyua, J., et al. (2020b). Genetic dissection of resistance to gray leaf spot by combining genome-wide association, linkage mapping, and genomic prediction in tropical maize germplasm. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.572027

Kibe, M., Nyaga, C., Nair, S. K., Beyene, Y., Das, B., Bright, J. M., et al. (2020a). Combination of linkage mapping, GWAS, and GP to dissect the genetic basis of common rust resistance in tropical maize germplasm. *Int. J. Mol. Sci.* 21, 6518. doi: 10.3390/ijms21186518

Kinyua, Z., Smith, J., Kibata, G., Simons, S., and Langat, B. (2011). Status of grey leaf spot disease in Kenyan maize production ecosystems. *Afr. Crop Sci. J.* 18 (4). doi: 10.4314/acsj.v18i4.68647

Knapp, S. J., Stroup, W. W., and Ross, W. M. (1985). Exact confidence intervals for heritability on a progeny mean basis. *Crop Science.* 25 (1), 192–194. doi: 10.2135/cropsci1985.0011183X002500010046x

Koo, Y. J., Kim, M. A., Kim, E. H., Song, J. T., Jung, C., Moon, J. K., et al. (2007). Overexpression of salicylic acid carboxyl methyltransferase reduces salicylic acid-mediated pathogen resistance in *Arabidopsis thaliana*. *Plant Mol. Biol.* 64, 1–15. doi: 10.1007/s11103-006-9123-x

Kuki, M. C., Pinto, R. J. B., Bertagna, F. A. B., Tessmann, D. J., Teixeira Do Amaral, A., Scapim, C. A., et al. (2020). Association mapping and genomic prediction for ear rot disease caused by Fusarium verticillioides in a tropical maize germplasm. *Crop Sci.* 60, 2867–2881. doi: 10.1002/csc2.20272

Kuki, M. C., Scapim, C. A., Rossi, E. S., Mangolin, C. A., Amaral Júnior, A. T. D., and Pinto, R. J. B. (2018). Genome wide association study for gray leaf spot resistance in tropical maize core. *PloS One* 13, e0199539. doi: 10.1371/journal.pone.0199539

Latterell, F. M. (1983). Gray leaf spot of corn: a disease on the move. *Plant Dis.* 67, 842. doi: 10.1094/PD-67-842

Lehmensiek, A., Esterhuizen, A. M., Van Staden, D., Nelson, S. W., and Retief, A. E. (2001). Genetic mapping of gray leaf spot (GLS) resistance genes in maize. *Theor. Appl. Genet.* 103, 797–803. doi: 10.1007/s001220100599

Lennon, J. R., Krakowsky, M., Goodman, M., Flint-Garcia, S., and Balint-Kurti, P. J. (2016). Identification of alleles conferring resistance to gray leaf spot in maize derived from its wild progenitor species teosinte. *Crop Sci.* 56, 209–218. doi: 10.2135/cropsci2014.07.0468

Li, G., Dong, Y., Zhao, Y., Tian, X., Würschum, T., Xue, J., et al. (2020). Genome-wide prediction in a hybrid maize population adapted to Northwest China. *Crop J.* 8, 830–842. doi: 10.1016/j.cj.2020.04.006

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

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, K.-J., and Xu, X.-D. (2013). First report of gray leaf spot of maize caused by cercospora zeina in China. *Plant Dis.* 97, 1656–1656. doi: 10.1094/PDIS-03-13-0280-PDN

Liu, L., Zhang, Y. D., Li, H. Y., Bi, Y. Q., Yu, L. J., Fan, X. M., et al. (2016). QTL mapping for gray leaf spot resistance in a tropical maize population. *Plant Dis.* 100, 304–312. doi: 10.1094/PDIS-08-14-0825-RE

Mahuku, G., Chen, J., Shrestha, R., Narro, L. A., Guerrero, K. V. O., Arcos, A. L., et al. (2016). Combined linkage and association mapping identifies a major QTL (Qrtsc8-1), conferring tar spot complex resistance in maize. *Theor. Appl. Genet.* 129, 1217–1229. doi: 10.1007/s00122-016-2698-y

Mammadov, J., Sun, X., Gao, Y., Ochsenfeld, C., Bakker, E., Ren, R., et al. (2015). Combining powers of linkage and association mapping for precise dissection of QTL controlling resistance to gray leaf spot disease in maize (Zea mays L.). *BMC Genomics* 16, 916. doi: 10.1186/s12864-015-2171-3

Martins, L. B., Rucker, E., Thomason, W., Wisser, R. J., Holland, J. B., and Balint-Kurti, P. (2019). Validation and characterization of maize multiple disease resistance QTL. *G3 Genes|Genomes|Genetics* 9, 2905–2912. doi: 10.1534/g3.119.400195

Meisel, B., Korsman, J., Kloppers, F. J., and Berger, D. K. (2009). Cercospora zeina is the causal agent of grey leaf spot disease of maize in southern Africa. *Eur. J. Plant Pathol.* 124, 577–583. doi: 10.1007/s10658-009-9443-1

Menkir, A., and Ayodele, M. (2005). [amp]]lsquo;Genetic analysis of resistance to gray leaf spot of midaltitude maize inbred lines'. *Crop Sci.* 45, 1. doi: 10.2135/cropsci2005.0163a

Meuwissen, T. H. E., Hayes, B. J., and Goddard, M. E. (2001). [amp]]lsquo;Prediction of total genetic value using genome-wide dense marker maps'. *Genetics* 157, 1819–1829. doi: 10.1093/genetics/157.4.1819

Nair, S. K., Babu, R., Magorokosho, C., Mahuku, G., Semagn, K., Beyene, Y., et al. (2015). Fine mapping of Msv1, a major QTL for resistance to Maize Streak Virus leads to development of production markers for breeding pipelines. *Theor. Appl. Genet.* 128, 1839–1854. doi: 10.1007/s00122-015-2551-8

Ndlovu, N., Kachapur, R. M., Beyene, Y., Das, B., Ogugo, V., Makumbi, D., et al. (2024). Linkage mapping and genomic prediction of grain quality traits in tropical maize (*Zea mays* L.). *Front. Genet.* 15. doi: 10.3389/fgene.2024.1353289

Ndlovu, N., Spillane, C., McKeown, P. C., Cairns, J. E., Das, B., and Gowda, M. (2022). Genome-wide association studies of grain yield and quality traits under optimum and low-nitrogen stress in tropical maize (Zea mays L.). *Theor. Appl. Genet.* 135, 4351–4370. doi: 10.1007/s00122-022-04224-7

Nyaga, C., Gowda, M., Beyene, Y., Muriithi, W. T., Makumbi, D., Olsen, M. S., et al. (2020). Genome-wide analyses and prediction of resistance to MLN in large tropical maize germplasm. *Genes* 11, 16. doi: 10.3390/genes11010016

Nzuve, N. (2013). Combining abilities of maize inbred lines for grey leaf spot (GLS), grain yield and selected agronomic traits in Kenya. *J. Plant Breed. Crop Sci.* 5, 41–47. doi: 10.5897/jpbcs12.056

Omondi, D. O., Dida, M. M., Berger, D. K., Beyene, Y., Nsibo, D. L., Juma, C., et al. (2023). Combination of linkage and association mapping with genomic prediction to infer QTL regions associated with gray leaf spot and northern corn leaf blight resistance in tropical maize. *Front. Genet.* 14. doi: 10.3389/fgene.2023.1282673

Poland, J. A., Bradbury, P. J., Buckler, E. S., and Nelson, R. J. (2011). Genome-wide nested association mapping of quantitative resistance to northern leaf blight in maize. *Proc. Natl. Acad. Sci. U.S.A.* 108, 6893–6898. doi: 10.1073/pnas.1010894108

Pozar, G., Butruille, D., Silva, H. D., McCuddin, Z. P., and Penna, J. C. V. (2009). Mapping and validation of quantitative trait loci for resistance to Cercospora zeae-maydis infection in tropical maize (*Zea mays L.*). *Theor. Appl. Genet.* 118, 553–564. doi: 10.1007/s00122-008-0920-2

Prasanna, B. M., Cairns, J. E., Zaidi, P. H., Beyene, Y., Makumbi, D., Gowda, M., et al. (2021). Beat the stress: breeding for climate resilience in maize for the tropical rainfed environments. *Theor. Appl. Genet.* 134, 1729–1752. doi: 10.1007/s00122-021-03773-7

Prasanna, B. M., Nair, S. K., Babu, R., Gowda, M., Zhang, X., Xu, Y., et al. (2020). "Increasing genetic Gains in Maize in Stress-Prone Environments of the Tropics," in *Genomic Designing of Climate-Smart Cereal Crops*. Ed. C. Kole (Springer International Publishing, Cham), 97–132. doi: 10.1007/978-3-319-93381-8_3

Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959.

Qiu, H., Li, C., Yang, W., Tan, K., Yi, Q., Yang, M., et al. (2021). Fine mapping of a new major QTL-qGLS8 for gray leaf spot resistance in maize. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.743869

Rashid, Z., Sofi, M., Harlapur, S. I., Kachapur, R. M., Dar, Z. A., Singh, P. K., et al. (2020). Genome-wide association studies in tropical maize germplasm reveal novel and

known genomic regions for resistance to Northern corn leaf blight. *Sci. Rep.* 10, 21949. doi: 10.1038/s41598-020-78928-5

Remington, D. L., Thornsberry, J. M., Matsuoka, Y., Wilson, L. M., Whitt, S. R., Doebley, J., et al. (2001). Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proc. Natl. Acad. Sci. USA* 98, 11479–11484. doi: 10.1073/pnas.201394398

Romay, M. C., Millard, M. J., Glaubitz, J. C., Peiffer, J. A., Swarts, K. L., Casstevens, T. M., et al. (2013). Comprehensive genotyping of the USA national maize inbred seed bank. *Genome Biol.* 14, R55. doi: 10.1186/gb-2013-14-6-r55

Sadessa, K., Beyene, Y., Ifie, B. E., Suresh, L. M., Olsen, M. S., Ogugo, V., et al. (2022). Identification of genomic regions associated with agronomic and disease resistance traits in a large set of multiple DH populations. *Genes* 13, 351. doi: 10.3390/genes13020351

Saghai Maroof, M. A. S., Yue, Y. G., Xiang, Z. X., Stromberg, E. L., and Rufener, G. K. (1996). Identification of quantitative trait loci controlling resistance to gray leaf spot disease in maize. *Theoret. Appl. Genet.* 93, 539–546. doi: 10.1007/BF00417945

Schnable, P. S., Ware, D., Fulton, R. S., Stein, J. C., Wei, F., Pasternak, S., et al. (2009). The B73 maize genome: complexity, diversity, and dynamics. *Science* 326, 1112–1115. doi: 10.1126/science.1178534

Schrag, T. A., Schipprack, W., and Melchinger, A. E. (2019). Across-years prediction of hybrid performance in maize using genomics. *Theor. Appl. Genet.* 132, 933–946. doi: 10.1007/s00122-018-3249-5

Shaner, G., and Finney, R. E. (1977). The effect of nitrogen fertilization in the expression of slow-mildewing resistance in Knox wheat. *Phytopathology* 67, 1051–1056. doi: 10.1094/Phyto-67-1051

Shi, L., Li, X., Hao, Z., Xie, C., Ji, H., Lü, X., et al. (2007). Comparative qtl mapping of resistance to gray leaf spot in maize based on bioinformatics. *Agric. Sci. China* 6, 1411–1419. doi: 10.1016/S1671-2927(08)60002-4

Shi, L., Lv, X., Weng, J., Zhu, H., Liu, C., Hao, Z., et al. (2014). Genetic characterization and linkage disequilibrium mapping of resistance to gray leaf spot in maize (*Zea mays L.*). *Crop J.* 2, 132–143. doi: 10.1016/j.cj.2014.02.001

Sitonik, C., Suresh, L. M., Beyene, Y., Olsen, M. S., Makumbi, D., Oliver, K., et al. (2019). Genetic architecture of maize chlorotic mottle virus and maize lethal necrosis through GWAS, linkage analysis and genomic prediction in tropical maize germplasm. *Theor. Appl. Genet.* 132, 2381–2399. doi: 10.1007/s00122-019-03360-x

Slaten, M. L., Chan, Y. O., Shrestha, V., Lipka, A. E., and Angelovici, R. (2020). HAPPI GWAS: holistic analysis with pre- and post-integration GWAS. *Bioinformatics* 36, 4655–4657. doi: 10.1093/bioinformatics/btaa589

Sserumaga, J. P., Makumbi, D., Assanga, S. O., Mageto, E. K., Njeri, S. G., Jumbo, B. M., et al. (2020). Identification and diversity of tropical maize inbred lines with resistance to common rust (*Puccinia sorghi* Schwein*). Crop Sci.* 60, 2971–2989. doi: 10.1002/csc2.20345

Tao, Y., Jiang, L., Liu, Q., Zhang, Y., Zhang, R., Ingvardsen, C. R., et al. (2013). Combined linkage and association mapping reveals candidates for Scmv1, a major locus involved in resistance to sugarcane mosaic virus (SCMV) in maize. *BMC Plant Biol.* 13, 162. doi: 10.1186/1471-2229-13-162

Technow, F., Bürger, A., and Melchinger, A. E. (2013). Genomic prediction of northern corn leaf blight resistance in maize with combined or separated training sets for heterotic groups. *G3 Genes|Genomes|Genetics* 3, 197–203. doi: 10.1534/g3.112.004630

Vivek, B. S., Odongo, O., Njuguna, J., Imanywoha, J., Bigirwa, G., Diallo, A., et al. (2010). Diallel analysis of grain yield and resistance to seven diseases of 12 African maize (*Zea mays* L.) inbred lines. *Euphytica* 172, 329–340. doi: 10.1007/s10681-009-9993-5

Wang, M., Yan, J., Zhao, J., Song, W., Zhang, X., Xiao, Y., et al. (2012). Genome-wide association study (GWAS) of resistance to head smut in maize. *Plant Sci.* 196, 125–131. doi: 10.1016/j.plantsci.2012.08.004

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

Warburton, M. L., Ribaut, J. M., Franco, J., Crossa, J., Dubreuil, P., and Betrán, F. J. (2005). Genetic characterization of 218 elite CIMMYT maize inbred lines using RFLP markers. *Euphytica* 142, 97–106. doi: 10.1007/s10681-005-0817-y

Ward, J. M. J., Stromberg, E. L., Nowell, D. C., and Nutter, F. W. (1999). Gray leaf spot: A disease of global importance in maize production. *Plant Dis.* 83, 884–895. doi: 10.1094/PDIS.1999.83.10.884

Wisniewski, M., and Rawlings, J. O. (1990). Applied regression analysis: A research tool. *J. Operational Res. Soc.* 41, 782–783. doi: 10.1057/jors.1990.106

Wu, F., Liu, Z., Xu, J., Gao, S., Lin, H., Liu, L., et al. (2016). Molecular evolution and association of natural variation in *zmARF31* with low phosphorus tolerance in maize. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.01076

Xu, L., Zhang, Y., Shao, S., Chen, W., Tan, J., Zhu, M., et al. (2014). High-resolution mapping and characterization of *qRgls2*, a major quantitative trait locus involved in maize resistance to gray leaf spot. *BMC Plant Biol.* 14, 230. doi: 10.1186/s12870-014-0230-6

Yigrem, M., and Yohannes, M. (2019). Participatory evaluation of maize varieties for the management of maize gray leaf spot disease (*Cercospora zeae maydis*) at Gondar Zuria district, northwest Ethiopia. *J. Dryland Agric.* 5 (1), 1–6. doi: 10.5897/joda2018.0001

Yu, J., and Buckler, E. S. (2006). Genetic association mapping and genome organization of maize. *Curr. Opin. Biotechnol.* 17, 155–160. doi: 10.1016/j.copbio.2006.02.003

Zarreen, F., Kumar, K., and Chakraborty, S. (2023). Phosphoinositides in plant-pathogen interaction: trends and perspectives. *Stress Biol.* 3, 4. doi: 10.1007/s44154-023-00082-5

Zhang, Y., Xu, L., Fan, X., Tan, J., Chen, W., and Xu, M. (2012). QTL mapping of resistance to gray leaf spot in maize. *Theor. Appl. Genet.* 125, 1797–1808. doi: 10.1007/s00122-012-1954-z

Zhang, X., Yang, Q., Rucker, E., Thomason, W., and Balint-Kurti, P. (2017). Fine mapping of a quantitative resistance gene for gray leaf spot of maize (Zea mays L.) derived from teosinte (*Z. mays ssp. parviglumis*). *Theor. Appl. Genet.* 130, 1285–1295. doi: 10.1007/s00122-017-2888-2

Zhao, Y., Gowda, M., Liu, W., Würschum, T., Maurer, H. P., Longin, F. H., et al. (2012). Accuracy of genomic selection in European maize elite breeding populations. *Theor. Appl. Genet.* 124, 769–776. doi: 10.1007/s00122-011-1745-y

Zheng, H., Chen, J., Mu, C., Makumbi, D., Xu, Y., and Mahuku, G. (2018). Combined linkage and association mapping reveal QTL for host plant resistance to common rust (*Puccinia sorghi*) in tropical maize. *BMC Plant Biol.* 18, 310. doi: 10.1186/s12870-018-1520-1

Zhu, M., Tong, L., Xu, M., and Zhong, T. (2021). Genetic dissection of maize disease resistance and its applications in molecular breeding. *Mol. Breed.* 41, 32. doi: 10.1007/s11032-021-01219-y

Zwonitzer, J. C., Coles, N. D., Krakowsky, M. D., Arellano, C., Holland, J. B., McMullen, M. D., et al. (2010). Mapping resistance quantitative trait loci for three foliar diseases in a maize recombinant inbred line population—Evidence for multiple disease resistance? *Phytopathology®* 100, 72–79. doi: 10.1094/PHYTO-100-1-0072

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 15 April 2025
DOI 10.3389/fpls.2025.1539641

Check for updates

OPEN ACCESS

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

REVIEWED BY
Asif Bashir Shikari,
Sher-e-Kashmir University of Agricultural Sciences and Technology of Kashmir, India
Ashish Kumar,
Jawaharlal Nehru Agricultural University, India

*CORRESPONDENCE
Somashekhar M. Punnuri
punnuris@fvsu.edu
Mahendar Thudi
mahendar.thudi@unisq.edu.au

RECEIVED 04 December 2024
ACCEPTED 27 February 2025
PUBLISHED 15 April 2025

CITATION
Sahu A, Rangari SK, Naik YD, Jyotish A, Pandey MK, Varshney RK, Thudi M and Punnuri SM (2025) Consensus genomic regions and key genes for biotic, abiotic and key nutritional traits identified using meta-QTL analysis in peanut.
Front. Plant Sci. 16:1539641.
doi: 10.3389/fpls.2025.1539641

COPYRIGHT
© 2025 Sahu, Rangari, Naik, Jyotish, Pandey, Varshney, Thudi and Punnuri. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Consensus genomic regions and key genes for biotic, abiotic and key nutritional traits identified using meta- QTL analysis in peanut

Aakash Sahu[1], Sagar Krushnaji Rangari[1,2], Yogesh Dashrath Naik[1], Anjali Jyotish[2], Manish K. Pandey[2], Rajeev K. Varshney[3], Mahendar Thudi[4,5]* and Somashekhar M. Punnuri[4]*

[1]Department of Agricultural Biotechnology and Molecular Biology, Dr. Rajendra Prasad Central Agricultural University (RPCAU), Pusa, Bihar, India, [2]Center of Excellence in Genomics & Systems Biology, International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Hyderabad, Telangana, India, [3]Western Australian (WA) State Agricultural Biotechnology Centre, Centre for Crop and Food Innovation, Murdoch University, Murdoch, WA, Australia, [4]College of Agriculture, Family Sciences and Technology, 1005 State University Dr, Fort Valley State University (FVSU), Fort Valley, GA, United States, [5]Centre for Crop Health, University of Southern Queensland (USQ), Toowoomba, QLD, Australia

Peanut (*Arachis hypogaea* L.), a key oilseed crop in the U.S., plays a significant role in agriculture and the economy but faces challenges from biotic and abiotic stresses, including aflatoxin contamination caused by *Aspergillus flavus* and *A. parasiticus*. Despite many large-effect QTLs identified for yield and key traits, their use in breeding is limited by unfavorable genetic interactions. To overcome this, we aimed to identify consensus genomic regions and candidate genes linked to key traits by analyzing QTL data from 30 independent studies conducted over the past 12 years, focusing on biotic, abiotic, aflatoxin, morphological, nutritional, phenological, and yield-associated traits. Using genetic map information, we constructed consensus maps and performed a meta-analysis on 891 QTLs, leading to the identification of 70 Meta-QTLs (MQTLs) with confidence intervals ranging from 0.07 to 9.63 cM and an average of 2.33 cM. This reduction in confidence intervals enhances the precision of trait mapping, making the identified MQTLs more applicable for breeding purposes. Furthermore, we identified key genes associated with aflatoxin resistance in MQTL5.2 (serine/threonine-protein kinase, BOI-related E3 ubiquitin-protein ligase), MQTL5.3, MQTL7.3, and MQTL13.1. Similarly, for yield-related traits in MQTL3.1–MQTL3.4 (mitogen-activated protein kinase, auxin response factor), MQTL11.2 (MADS-box protein, squamosa promoter-binding protein), and MQTL14.1. Genes related to oil composition within MQTL5.2 (fatty-acid desaturase FAD2, linoleate 9S-lipoxygenase), MQTL9.3, MQTL19.1 (acyl-CoA-binding protein, fatty acyl-CoA reductase FAR1), MQTL19.4, and MQTL19.5. Nutritional traits like iron and zinc content are linked to MQTL1.1 (probable methyltransferase, ferredoxin C), MQTL10.1, and MQTL12.1. These regions and genes serve as precise targets for marker-assisted breeding to enhance peanut yield, resilience, and quality.

KEYWORDS

peanut, aflatoxin, biotic and abiotic stresses, meta-QTL, genes, yield

1 Introduction

Peanut (*Arachis hypogaea*), also known as groundnut or earthnut, is an allotetraploid oilseed legume crop (2n=4x=40) with a genome size of 2.7 Gb. It is the second most widely cultivated legume crop globally, primarily grown for oil production, following soybean (*Glycine max* L.; Vishwakarma et al., 2017). Peanuts are a key ingredient in ready-to-use therapeutic and supplementary foods aimed at combating malnutrition in underdeveloped and developing regions (Variath and Janila, 2017). Peanut seeds contain approximately 45–50% oil, 25% protein, 15% carbohydrates, and various beneficial secondary compounds. They also provide all 20 amino acids, making a significant contribution to human nutrition (Shasidhar et al., 2017; Wang et al., 2015).

Global peanut production reached 539 lakh tons, harvested from 327 lakh hectares, with an average productivity of 1,648 kg per hectare (FAOSTAT, 2021). China, India, Nigeria, the United States, and Sudan together account for about 69% of global peanut production (USDA, 2024; https://ipad.fas.usda.gov/cropexplorer/cropview/commodityView.aspx?cropid=2221000). In 2022–2023, India exported approximately 668,885.40 metric tons of peanuts (APEDA, 2024; https://apeda.gov.in/apedawebsite/GroundNut/GroundNut.htm). Notably, around 70% of the peanut cultivation area is located in arid and semi-arid regions. Peanuts are highly adaptable, thriving in soils with low chemical inputs, requiring minimal water, and exhibiting the lowest carbon footprint among nuts (Pandey et al., 2020). Additionally, peanut plants are zero-waste crops and play a role in preventing soil erosion (National Peanut Board, 2023; Agropedia, 2023).

Despite their significant economic and nutritional importance, peanut production faces substantial challenges from biotic and abiotic stresses, including pests, diseases, and adverse environmental conditions. One major constraint is aflatoxin contamination, which severely reduces peanut quality and poses serious health risks to humans (Pandey et al., 2017). Additionally, diseases such as tomato spotted wilt virus (TSWV), early leaf spot (ELS), late leaf spot (LLS), rust, root-knot nematode (RKN), and stem rot (SR) are critical biotic threats to peanut production worldwide. These challenges lead to significant yield losses and adversely affect the quality of peanut oil and seeds.

While both high and low oil content in peanuts have value—small seeds are preferred for oil production and large seeds for confectionery products (Wang et al., 2015; Gangurde et al., 2022)—a high oleic-to-linoleic acid ratio is particularly desirable. This trait improves shelf life and provides health benefits for both manufacturers and consumers (Wang et al., 2015; Pandey et al., 2014). Farmers and traders can benefit substantially from incorporating stable, high-oil content traits into elite cultivars, as even a 1% increase in oil content can raise producer profits by 7% (Liao, 2003). Many agronomic traits in peanuts are inherited quantitatively and influenced by genotype × environment interactions, highlighting the importance of identifying stable and promising genomic regions for crop improvement. Discovering such genomic regions, especially those conferring resistance to aflatoxins, is critical for sustainable peanut production (Yu et al., 2019; Variath and Janila, 2017).

The primary focus of peanut breeding is to improve yield, which is closely linked to pod, seed, and disease-related traits. However, these are complex quantitative traits that can be enhanced through quantitative trait loci (QTL) mapping (Lu et al., 2018). Over the two decades, numerous QTL studies on important traits have been conducted (PeanutBase, 2023; Ahmad et al., 2020; Dodia, 2018; Lu et al., 2018; Pandey et al., 2017; Bertioli et al., 2016; Sujay et al., 2012). However, many of these studies are limited to specific environments, leading to inconsistent and unstable phenotypes when these traits are introduced into elite cultivars with new genetic backgrounds. This instability often arises from unfavorable epistatic interactions (Lu et al., 2018). Consensus genomic regions and candidate genes play a pivotal role in the efficient transfer of desirable traits into crops, significantly reducing the risk of genetic load. Genetic load refers to the accumulation of unfavorable or non-beneficial alleles that can arise during traditional breeding approaches. By focusing on consensus genomic regions identified through meta-QTL (MQTL) analysis, breeders can target stable and well-defined loci associated with important traits, ensuring precision and efficiency in marker-assisted selection (MAS).

MQTL analysis has emerged as a powerful approach to refining and consolidating data from multiple QTL studies, thereby improving the precision and utility of genetic markers for breeding (Kumar et al., 2023; Isha et al., 2024; Gupta et al., 2024). While individual QTL studies provide valuable insights, they often report QTLs with large confidence intervals (CI) and inconsistencies across different environments and genetic backgrounds. MQTL analysis addresses these limitations by integrating data from multiple independent QTL studies to identify consensus QTLs with smaller CI, thereby enhancing the resolution of trait-associated regions. In this study, we report MQTLs for various biotic and abiotic stress resistance traits and key nutritional traits in peanuts, based on 30 QTL studies available in the public domain up to 2024.

2 Material and methods

2.1 Collection of studies and QTLs

A comprehensive review was conducted to gather information on QTLs associated with abiotic stress, biotic stress, morphological, nutritional, phenological, physiological, and yield-related traits in peanut over the past 12 years. A total of 30 studies were sourced from platforms such as Google Scholar and PubMed. Key details, including population type, logarithm of odds (LOD) score, phenotypic variation explained (PVE/R^2), population size, marker positions, QTL locations, and CIs, were extracted from these research articles. Each QTL was treated as an individual entity, even if it was identified in multiple environments or genetic backgrounds. In total, 893 QTLs were initially utilized to conduct the MQTL analysis (Supplementary Table S1). All these QTLs were grouped into eight major trait categories (1) aflatoxin (percent seed infection index PSIIA, PSIIB, RAF, AFB2A, AFTA and AFTB), (2)

abiotic stress-associated traits (DW, HW, SLA, TR, TE, SDW and ISC), (3) biotic stress-associated traits (AE, ELS, LLS, LS, NH, rust, stem rot, thrips, TSWV, root knot nematodes, NH and bruchid), (4) morphological traits (NB, HRN, IN, NN, PEL, NPB, NSB, IL, PS, TLW, PH, PC, PEL, MSH and PL), (5) nutritional traits (Fe, Zn, OC, OA, AA, BA, GA, LGA, LA, OC, OLR, PA and SA), (6) phenological traits (DE, DF, EDP, TDP, DFF, FI and DW), (7) physiological traits (SCMR, CT, HUE, LDW, NID, SD, LDW, CID, SC, SLW, VCR, RWC and LA), (8) yield related traits (10-SW, HSW, BDW, DPN, HI, PL, DPL, PN, PWE, PWL, PW, PWD, DPW, SN, SWT, SW, SP, SDW, SP and LW). To calculate start CI and end CI positions, we used the following formula for recombinant inbred lines (RIL), back cross (BC) and F_2 populations (Darvasi and Soller, 1997; Venske et al., 2019; Guo et al., 2006).

For RIL populations, CI = $\frac{163}{P*R2}$ and for F_2 and BC populations, CI = $\frac{530}{P*R2}$ Where P refers to the size of the population and R^2 refers to the phenotypic variance explained. The peak, initial, and final positions of the QTLs are also determined for the QTL projection and the MQTL analysis.

2.2 Construction of high-density consensus maps

High density consensus maps for all 20 chromosomes were constructed with the help of 'LPmerge' package of R software. Markers and their positions from each map of each linkage group were extracted, and '.csv' extension format Excel sheets were prepared as an input file in the R studio. The commands for the LPmerge package were modified according to input data (Endelman and Plomion, 2014). The LPmerge package uses linear programming to efficiently minimize the mean absolute error between the consensus maps and the linkage maps from each population. To obtain the weighted consensus maps, population size of each map was provided in the commands. Then, it creates the "n" number of weighted models, which is selected based on the smallest length of the consensus map and root mean sum of square error (RMSE) value. The best consensus map for all the 20 linkage groups was saved in the '.csv' extension file format.

2.3 QTL projection on the consensus maps

The consensus maps constructed using LPmerge was used for MQTL analysis using Biomercator software (Arcade et al., 2004; Olivier et al., 2012; https://mybiosoftware.com/biomercator-genetic-maps-qtl-integration.html). In addition to the consensus map file, QTL information files were also created in '.txt' format. The Veyrieras two-step algorithm (Veyrieras et al., 2007), included in the BioMercator v4.2 software, was used to perform meta-analysis. In the first step, meta-analysis determines the best MQTL model based on model choice criteria from the Akaike information criterion (AIC), a corrected AIC, a Bayesian information criterion (BIC) and the average weight of evidence (AWE). The best MQTL model was selected based on the lowest value and highest weight. All the output information regarding MQTLs and their QTLs from all the 20 linkage groups were extracted. The MQTLs that correspond to weak associations were excluded. MQTL regions containing more than two QTLs for different traits and from different studies were only considered for further analysis. The QTL regions, which deviated from their MQTL positions, were also excluded from the MQTL group. QTL nomenclature is as follows: the name starts with MQTL, followed by the number of the consensus map where the QTL was detected, and a serial number for those MQTLs where two or more were found.

2.4 Candidate genes identification from the MQTL region

We used the physical positions of flanking markers from the MQTL regions to retrieve genes from the "NCBI" database for candidate gene identifications. In cases where the physical positions of the flanking markers were not available, the positions of adjacent markers were used instead. Genes were identified using the reference genome assembly of peanut (arahy.Tifrunner.gnm1.KYV3). Genes located in these regions with established functional links to the trait of interest in any crop species were regarded as potential candidate genes for further analysis.

3 Results

3.1 Salient features of the QTL studies

Thirty independent studies used different types of mapping populations, such as RILs, BC, and F_2. A total of 893 QTLs were used for the meta-analysis, and all were classified into different categories such as aflatoxin, biotic and abiotic stress-associated traits, morphological, phenological and yield-associated traits (Supplementary Table S1). Of these, 30 QTLs related to aflatoxin, 57 QTLs related to abiotic stresses, 158 QTLs related to biotic stresses, 61 QTLs related to morphological traits, 200 QTLs related to nutritional traits, 50 QTLs related to phenological traits, and 87 QTLs related to physiological traits, 250 QTLs related to yield were used in this study. The maximum number of QTLs were observed on LG05 (Linkage Group 05) with 74 QTLs, while the minimum 22 QTLs were on LG17 (Supplementary Table S2). The phenotypic variation explained (PVE) varied from 0.013 to 91.1%. Out of 893 QTLs, 437 major QTLs (PVE ≥ 10%) and 456 minor QTLs (PVE ≤10%) were selected for MQTL analysis (Supplementary Table S1).

3.2 Construction of consensus maps

The consensus map was constructed utilizing pre-existing linkage map data, resulting in a comprehensive map with 11,956 markers spanning 4,496.20 cM (Supplementary Table S3). This map contains various types of molecular markers, such as SSRs and

SNPs. The distribution of markers per linkage group also showed considerable variation, ranging from 340 markers on LG10 to 1,006 markers on LG05. On an average 2.76 markers per cM were mapped on the consensus genetic map constructed. Linkage group 10 was the shortest, with a length of 128.05 cM, and had the least number of markers. Linkage group 13 was the longest, with a length of 337.9 cM. In addition, we observed substantial variability in marker density across linkage groups (Supplementary Table S3). For instance, LG05 has a high marker density of 4.73 markers per cM, while LG02 has a lower density of 1.55 markers per cM, indicating more densely mapped regions in specific chromosomes.

3.3 Meta-QTLs detection and their distribution on the peanut genome

In the present study, a total of 70 MQTLs were identified across 20 different LGs (Table 1). The distribution of MQTLs showed

TABLE 1 Meta-QTLs identified on 20 linkage groups of peanut for various traits.

MetaQTL	LG	Position	CI (95%)	CI start	CI end	Left Marker	Right Marker	No. of QTL	Traits
MQTL1.1	1	5.3	1.12	4.74	5.86	5939F	C39	5	TR, Fe, TSWV, SLA and Zn
MQTL1.2	1	56.33	1.78	55.44	57.22	Marker64	bgPabg-595666	8	SLA, PEL, KPA, SDW, PWPP, IN and SP
MQTL1.3	1	73.14	3.34	71.47	74.81	Ah01_38930228	Marker460	5	HUE, LLS, HSW, CT and PW
MQTL2.1	2	0.39	0.5	0.14	0.75	Marker927	Marker893	9	HSW, LW, SL and LW
MQTL2.2	2	15.48	3.58	13.69	5.37	Ah02_98843120	IPAHM524a	4	KDLL, CT and AE
MQTL2.3	2	53.94	1.12	53.38	1.68	GM2196-900	Seq11G07	8	SLA, SP, TW, PYPP, NPPP and PWPP
MQTL2.4	2	207.80	0.28	207.66	0.42	RGC2	Ah5600	8	YOC, PWPP, TDP, HI, TR and HW
MQTL3.1	3	0.35	1.44	0.37	1.07	bgPabg-596210	AX-147243049	8	NPB, PYPP, DFF, SLW, KPA and HSW
MQTL3.2	3	12.03	1.61	11.22	12.83	ARS761-300	PM477	4	SDW, Zn, LS and Fe
MQTL3.3	3	69.72	0.07	69.68	69.75	AX-147217292	AX-147217370	10	MPTO, Fe, WBA, OIL, LLS, NSB and ELS
MQTL3.4	3	81.3	3.53	79.53	83.06	GM2206	PM238-1	4	HUE, PW and HSW
MQTL3.5	3	115.46	2.22	114.35	116.57	AHGS1674	GM2691	3	SLA and SDW
MQTL3.6	3	129.03	2.02	128.02	130.04	AX-147217953	Ah03_133970796	5	NH, VCR and TW
MQTL3.7	3	151.24	0.32	151.08	151.4	Ah03_126798348	Seq15F12	4	SCMR, PW and SW
MQTL4.1	4	10.08	1.21	9.475	10.68	bgPt-593893	bgPt-600898	4	NPPP, TSW and VCR
MQTL4.2	4	34.33	2.61	33.02	35.63	bgPabg-597624	GM890	4	WPA, LLS and TSW
MQTL4.3	4	64.86	0.20	64.76	64.96	AX-147219426	Ah15_155617956	7	Fe, AE, TDP, ISC, PLB and HSW
MQTL5.1	5	15.37	1.20	14.77	15.97	Marker3583	Ah5897	4	LW, SL and KDLL
MQTL5.2	5	30.30	2.20	29.2	31.4	Ah6140	Ah6242	4	WSPA, SL, MPSO and SW
MQTL5.3	5	55.64	2.94	54.17	57.11	Ah4097	Ah5485	3	MPSO and ELS
MQTL5.4	5	88.57	1.44	87.85	89.29	AHGS2534	Ahs2641II	15	YAA, OIL, SP, LLS and PC
MQTL5.5	5	96.46	2.05	95.43	97.485	A05A1146	A05A1355	4	LLS, HPW, AE and PLA
MQTL5.6	5	103.93	1.92	102.97	104.89	AX-147250857	AX-147223064	5	HPW, RLLS, WA and PLA
MQTL5.7	5	131.36	2.86	129.93	132.79	qHW-A05.2	PM112	7	SLA, HW, TE and DW
MQTL5.8	5	140.43	0.19	140.33	140.52	Ah05_114999121	AX-147223487	8	TR, SLA, NB, HW, SDW and DW
MQTL6.1	6	14.31	3.22	12.7	15.92	Ah06_3810427	bgPabg-597436	3	EL, PWA and ELS
MQTL6.2	6	50.12	4.07	48.08	52.15	IPAHM509	Ah06_21266806	3	LLS, NPB and NSB
MQTL6.3	6	74.52	3.85	72.59	76.44	GNB837	TC1A08	4	CT, SP and SCMR

(Continued)

TABLE 1 Continued

MetaQTL	LG	Position	CI (95%)	CI start	CI end	Left Marker	Right Marker	No. of QTL	Traits
MQTL7.1	7	12.45	2.16	11.37	13.53	bgPabg-594537	seq3B8-400	10	PAA, KPA, SDW, TE, EL, Fe, WSS, SdWA and ISC
MQTL7.2	7	37.54	2.51	36.285	38.795	Marker4573	PM450	11	LDW, GDDFI, HW, CT, SLA, NPB, SDW, PW, SdW and FI
MQTL7.3	7	85.75	1.82	84.84	86.66	Ah6254	AX-147227736	10	WSA, SW, WSPA, NB, ISC, TE and SCMR
MQTL7.4	7	101.29	0.85	100.86	101.71	AHGA102053	GM1986-2	5	PW, DFF, SP and WSB
MQTL8.1	8	2.10	2.38	0.91	3.29	AX-147231702	PM367	4	KDLL, NH and KDrust
MQTL8.2	8	16.51	1.74	15.64	17.38	AX-147229678	Ah6347	5	YPA, DFF, KDrust, RWC and Zn
MQTL8.3	8	58.92	5.78	56.03	61.81	Ahs5872I	AX-147229847	4	YOC, NPPP, KDrust and YLGA
MQTL8.4	8	252.61	2.76	251.23	253.99	TC14B08	A08B47	5	WB, MPTLA, LLS, WPA and MPTO
MQTL9.1	9	3.61	2.06	2.58	4.64	S9_81536603	TC2D08	6	KDrust, TR and KDLL
MQTL9.2	9	84.58	2.12	83.52	85.64	Ah09_117710447	TC1D02	13	PTIFF, PLA, CT, WAA, PW, HUE, TSW and SP
MQTL9.3	9	104.27	1.22	103.66	104.88	Marker5532	AX-147234176	5	OIL, OLE, PAL and LIN
MQTL9.4	9	122.17	1.23	121.55	122.78	AhTE0572	FAD2A	11	MPTL, MPSL, MPTO, YOA, NSB and MPSO
MQTL9.5	9	178.79	0.69	178.44	179.13	Ah6234	Ah6116	5	WSA, SLA, YBA, YLG and YSA
MQTL10.1	10	6.8	2.30	5.66	7.94	TC7H11	Ah6326	6	Thrips, TDP, Kdrust, LS and YOC
MQTL10.2	10	71.98	0.38	71.79	72.16	AhTE0162	Marker6000	7	HW, SP, OLE, SIL, CT and PC
MQTL11.1	11	0.47	9.63	4.34	5.28	Marker6137	seq2G4-1	5	ISC, PLE, LS, TSW and KDLL
MQTL11.2	11	50.05	2.28	48.91	51.19	TOG896615_1198#	SD_c329p435vAC#	14	PW, ELS, DW, TR, SW, PC, HW, LPL and YLGA
MQTL11.3	11	117.84	3.01	116.33	119.34	PM83	PM52	4	SDW, Zn and EDP
MQTL12.1	12	20.54	7.61	16.735	24.34	AHS0046#	AHGS1692_b1	6	DPW, TSW, PLE, DPL and PWD
MQTL12.2	12	42.85	4.26	40.72	44.98	IPAHM531	A02B349	4	SDW, NPPP and MSH
MQTL12.3	12	128.56	2.67	127.22	129.89	Ah12_3356216	GNB1121	4	TW
MQTL13.1	13	61.47	1.11	60.915	62.02	AX-147216965	AX-147252574	9	PTIFF, KDLLS, SW, TE, VCR, HW, SCMR and Zn
MQTL13.2	13	133.4	2.85	131.97	134.82	AX-147217982	Ah2820	9	WSG, MPSO, YOC, Stem Rot, WSP, NPPP, WSL and OIL
MQTL13.3	13	256.94	5.34	254.27	259.61	AHGS1571	RN10F09	3	WG, SN and MSH
MQTL14.1	14	21.11	1.60	20.31	21.91	S14_22478715	GM1959-185	18	TE, SN, MSH, Rust, HW, 10SW, LS, TSW, NPPP, PC, MSH, Fe, Stem Rot and PL
MQTL14.2	14	89.78	3.28	88.14	91.42	AX-147247139	AX-147247229	5	LLS, HSW, MPSO, MPTO and HSW
MQTL14.3	14	112.64	0.08	112.6	112.68	AHGS0202	AX-147219990	9	MPSL, WL, WS, WA, WP, WB, WG, WSS and YPA
MQTL15.1	15	103.97	4.45	101.74	106.19	AX-147250857	Ah15_26507737	3	PW, DFF and FI
MQTL15.2	15	123.2	1.78	122.31	124.09	S1066EaB	AX-147223295	3	FI, PH and DFF
MQTL16.1	16	11.53	1.08	10.99	12.07	PM210	Marker9514	10	PC, HSW, SCMR, PC, SP and SW
MQTL16.2	16	38.66	4.93	36.19	41.12	AX-147226634	S16_74072826	3	ELS, SP and MPTO
MQTL16.3	16	66.21	6.22	63.1	69.32	Ah16_2247722	Ahs3386	3	WP, WA and Stem Rot

(Continued)

TABLE 1 Continued

MetaQTL	LG	Position	CI (95%)	CI start	CI end	Left Marker	Right Marker	No. of QTL	Traits
MQTL17.1	17	14.89	3.00	13.39	16.39	AX-147227003	AX-147254568	4	DPA, PW, KP and KPA
MQTL17.2	17	67.19	3.50	65.44	68.94	S17_65719159	S17_38129204	3	PL, SP and HSW
MQTL18	18	192.09	0.50	191.84	192.34	GNB159	GM1986-1	9	PTOLR, MPTO, MPTL, WP, MPSO and LLS
MQTL19.1	19	19.11	0.51	18.85	19.36	Marker11860	Ah19_11699918	44	OLA, DFF, PC, LIN, OLE, PL, ShW, CT, PLE and LA
MQTL19.2	19	90.51	0.57	90.22	90.79	FAD2B	AhS67426	10	YPA, MPSL, WSP, MPSLOA, YOC, MPSOLR and YLA
MQTL19.3	19	117.13	0.68	116.79	117.47	gi30420405	AX-147232613	3	YOC and HW
MQTL20.1	20	70.63	2.52	69.37	71.89	Ah20_117411384	AHGS1446	8	Stem Rot, OIL, TSW and TW
MQTL20.2	20	102.84	3.69	100.99	104.68	GM2165	AHGA75537	5	SN, NSB, SP, DW and GDDFI
MQTL20.3	20	116.03	2.03	115.01	117.04	Ah20_143925737	Ah20_126361289	8	HSW, PYPP, GDDFI, SLA, CT and SLA
MQTL20.4	20	136.76	0.69	136.41	137.10	Ah20_143912200	Ah20_134973352	5	HSW, ShW, SP and HI

10SW, Ten seed weight; AA, Arachidic acid; AFB2A, Resistance to production of aflatoxin B2; AE, Adult emergence; AFTA, Aflatoxin content; AFTB, Aflatoxin content; BA, Behenic acid; BC, Backcross; BDW, Biomass dry weight; CI, Confidence interval; CID, Leaf carbon isotope analysis; CT, Canopy temperature; DE, Days of emergence; DF, Days to flowering; DFF, Days to 50% flowering; DPL, Pod length of double seeded; DPW, Pod width of double seeded; DPN, Double seeded pod; DW, Dry weight; EDP, Estimated days to podding; ELS, Early leaf spot; Fe, Iron content; FI, Flower initiation; GA, Gadoleic acid; HI, Harvest index; HRN, Root hairiness; HSW, Hundred seed weight; HUE, Heat use efficiency; HW, Haulm weight; IL, Internode length; ISC, Delta biomass canopy conductance; LA, Leaf area; LDW, Leaf dry weight; LLS, Late leaf spot; LOD, Logarithm of odds; LW, Terminal leaflet width; MQTL, Meta QTL; MSH, Main stem height; NB, Number of branches; NID, Leaf isotope analysis; NH, Number of holes on pod; NN, Node number; NPB, Number of primary branches; NSB, Number of secondary branches; OA, Oleic acid; OC, Oil content; OLR, Oleic to linoleic acid ratio; PA, Palmitic acid; PC, Pod constriction; PEL, Peduncle length; PH, Plant height; PL, Pod length; PN, Pod number per plant; PS, Plant spread; PSIIB, Percent seed infection index for 100 seed weight; PSIIA, Percent seed infection index for 100 seed weight; PSII, Percent seed infection index; PW, Pod weight per plant; PWD, Pod width; PWE, Pod weight; PWL, Pod weight loss; QTLs, Quantitative trait locus; RAF, Aspergillus favus resistance; RILs, Recombinant inbred lines; RWC, Relative water content; SA, Stearic acid; SCMR, SPAD chlorophyll meter reading; SC, Stomatal conductance; SD, Stomatal density; SDW, Shoot dry weight; SDW, Weight of two kernels; SLA, Specific leaf area; SLW, Specific leaf weight; SN, Seed number; SP, Shelling percentage; SWT, Seed weight per plant; TE, Transpiration efficiency; TLW, Terminal leaflet width; TDP, Total developmental period; TR, Transpiration rate; TSWV, Tomato spotted wilt virus; VCR, Visual chlorotic rating; Zn, Zinc content; KPA, Kernel percentage; PWPP, Pod weight per plant; IN, Internode length; KDLL, Late leaf; spot; PYPP, Pod yield per plant; TW, Test weight; NPPP, No of pods per plant; YOC, Oil content; OIL, Oil content; MPTOA, Oleic acid; WBA, Behenic acid; WPA, Palmitic acid; PLB, Pod length; WSPA, Palmitic acid; MPSO, Oleic acid; YAA, Arachidic acid; HPW, Pod weight per plant; RLLS, Late leaf spot; WSS, Stearic acid; GDDFI, Days to flower initiation; WSA, Stearic acid; TE, Transpiration efficiency under WW regime; WSPA, Palmitic acid; WSB, Behenic acid; KDrust, Rust; YLGA, Gadoleic acid; WB, Behenic acid; PAL, Palmitic acid; WSA, Stearic acid; YBA, Behenic acid; YLG, Gadoleic acid; SP, Shelling percentage; SIL, Seed length; PLE, Pod length; LPL, Pod length; YLGA, Gadoleic acid; WSG, Gadoleic acid; WS, Stearic acid; WSS, Stearic acid; PTOLR; Oleic to linoleic acid ratio; MPTO, Oleic acid; MPTL, Linoleic acid; WP, Palmitic acid; MPSO, Oil content; LIN, Linoleic acid content; OLA, Oleic acid; OLE, Oleic acid; MPSLOA, Oleic acid; YOC, Oil content; MPSOLR, Oleic to linoleic acid ratio; ShW, Shell weight.

noticeable variability across these regions. LG05 harbored the highest number of MQTLs, with 8 (11.43% of the total), followed by LG03 with 7 MQTLs (10%), and LG09 with 5 MQTLs (7.14%; Figure 1). These regions are likely critical for important traits such as disease resistance and seed weight. Other linkage groups, such as LG02, LG07, LG08 and LG20, each contained 4 MQTLs (5.71%), while LG18 had the only single MQTLs (Supplementary Table S2). The uneven distribution of MQTLs across the genome points to specific genomic hotspots that may be associated with key traits. Moreover, certain MQTLs, such as MQTL5.4, were associated with many underlying QTLs (15 in this case), which are linked to several traits like late leaf spot (LLS), oil content, and seed size. This suggests that some MQTLs are hotspots for the genetic control of multiple traits, providing an invaluable resource for breeders aiming to improve multiple traits simultaneously. Among the 70 MQTLs, 13 MQTLs contained 18 aflatoxin-associated QTLs. Five aflatoxin QTLs were located on LG05 across three different MQTLs: MQTL5.1, MQTL5.2, and MQTL5.3. Additionally, two QTLs each were identified on LG07 (MQTL7.3), LG13 (MQTL13.1), and LG16 (16.2). MQTL5.2 contains six QTLs associated with palmitic acid, oleic acid, seed weight, seed length and two related to aflatoxin. Similarly, MQTL5.3 grouped two aflatoxin QTLs along with other quality and disease-associated QTLs. MQTL14.1 was crucial for both disease resistance and yield traits, underscoring its role in improving both plant health and productivity. MQTL9.7, MQTL19.1 and MQTL18.8 were significant for enhancing oil content and quality, focusing on the balance of linoleic and oleic acids. MQTL1.4 impacts various morphological traits, offering opportunities to optimize plant structure and pod characteristics. MQTL7.5 addresses physiological traits influencing factors like chlorophyll content and transpiration efficiency, which are vital for plant growth and stress resilience. MQTL10.1 harbors QTLs for multiple traits such as oil content, thrips resistance, leaf spot and rust resistance. Collectively, these MQTLs provide valuable insights for targeted breeding strategies aimed at improving peanut quality, yield, and resilience.

The precision and stability of the identified MQTLs are evident in their CIs. Initially, the average CIs of the individual QTLs ranged from 1.59 to 29.37 cM, with a mean of 9.26 cM, highlighting variability in the original QTL positions. After meta-analysis, the MQTL CIs significantly decreased, ranging from 0.07 cM (MQTL3.3) to 9.63 cM (MQTL11.1), with an average of 2.33 cM,

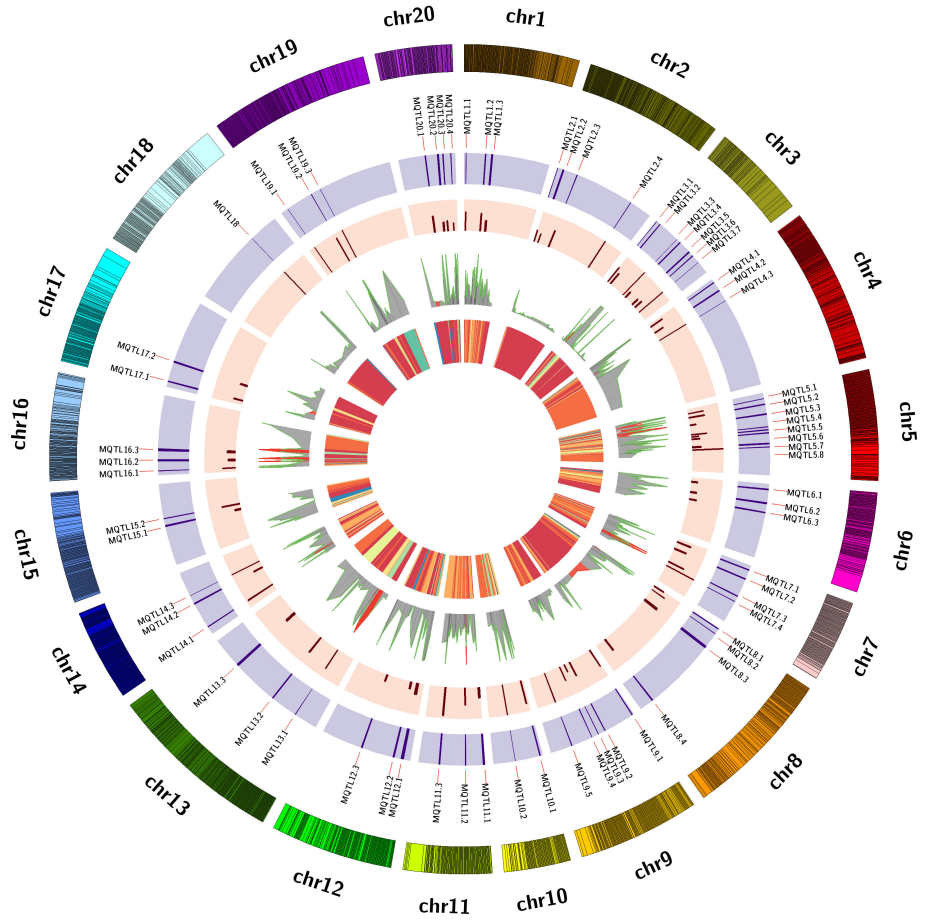


FIGURE 1
The 70 MQTLs for important traits across 20 chromosomes is illustrated. The circles from inside to outside represent the following: the 1. heatmap of QTLs, 2. projected QTLs with their PVE (Phenotypic Variance Explained), 3. fold change per Meta-QTL, 4. Meta-QTLs with their confidence Intervals (CI), and 5. chromosome wise marker density, respectively.

marking a notable improvement in precision. The fold change in CI, reflects enhanced localization accuracy, varied from 1.76 to 111.38, with an average fold change of 9.02, showcasing substantial improvement in MQTL precision over the original QTLs (Supplementary Table S4). Further analysis across linkage groups (LGs) revealed that the average CI of the original QTLs ranged from 5.06 to 17.73 cM, with an overall average of 9.23 cM, indicating considerable variability. After meta-analysis, MQTL CIs were reduced across all LGs, ranging from 0.50 to 4.97 cM, with an average of 2.40 cM (Supplementary Table S2; Figure 2), demonstrating increased precision across the genome. The fold change in CI reduction also varied across LGs, with LG18 showing the largest fold change (21.58), indicating a substantial improvement in QTL localization. LG10, LG14, and LG19 displayed high fold changes of 8.13, 7.04, and 8.72, respectively, further emphasizing the enhanced precision in these regions (Supplementary Table S2). In contrast, LG06 and LG16 had the lowest fold changes, 1.99 and 1.77, respectively, indicating more modest refinements. On average, the fold change across all LGs was 5.17, demonstrating an overall improvement in MQTL localization compared to the original QTLs.

3.4 Candidate genes identification

For candidate gene identification, MQTLs less than 50 Mb in size were used. Total 48 MQTL regions were selected for the identification of candidate genes based on the availability of the physical position of the flanking markers of MQTLs (Supplementary Table S5). From these meta-genomic regions, we identified a total of 4,024 genes (Supplementary Table S6). The functions of these genes were determined through a detailed review of the existing literature. MQTL1.1 is harbors QTLs for transpiration rate, iron content, tomato spotted wilt virus resistance, specific leaf area, and zinc content. For transpiration rate, genes like gibberellin-regulated protein 14-like (*LOC112710127*), caffeoyl-shikimate esterase-like (*LOC112720721*) and EPIDERMAL PATTERNING FACTOR-like protein (*LOC112754391*) were identified. Similarly, gibberellin-regulated protein (*LOC112710127*), transcription factor KUA1 (*LOC112802418*), and several small nuclear RNAs were found to be associated with specific leaf area (Miceli et al., 2019). Disease resistance related genes like serine/threonine-protein kinase (*LOC112801194*), BOI-related E3 ubiquitin-protein ligase (*LOC112801426*), autophagy-related protein (*LOC112802850*), auxin-responsive protein IAA30-like

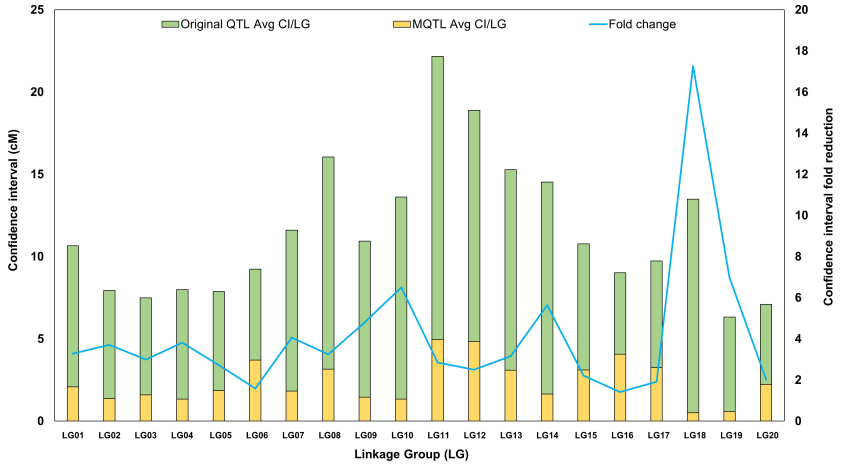


FIGURE 2
Graphical representation of projected QTLs and Meta-QTLs. Comparison of the mean confidence interval (CI) of original QTLs (green bar) and meta-QTLs (yellow bar). The blue line indicates the average fold reduction in Meta-QTLs.

(*LOC112718703*), WAT1-related protein (*LOC112802633*), and protein YABBY 4 (*LOC112803178*) were identified from this region. Iron and zinc content related genes are probable methyltransferases PMT14 (*LOC112801274*), FCS-Like Zinc finger 2 (*LOC112801380*), formyltetrahydrofolate deformylase 1 (*LOC112802537*), ferredoxin C 2 (*LOC112801048*), ion protease homolog 2, peroxisomal (*LOC112801369*), zinc finger A20 and AN1 domain-containing stress-associated protein (*LOC112803256*), transcription factor bHLH18 (*LOC112801016*). MQTL 1.7 is harbors QTLs for the leaf spot, late leaf spot, and rust resistance. We identified some of the important disease resistance related genes from this region like serine/threonine-protein kinase STY13 (*LOC112695569*), ubiquitin-conjugating enzyme E2-23 (*LOC112695673*), WAT1-related protein (*LOC112695708*). On chromosome 3, four MQTL regions were found, and most of the QTLs in this region are yield-related traits. The genes related to these traits are mitogen-activated protein kinase (*LOC112789588*), E3 ubiquitin-protein ligase PUB23 (*LOC112789537*), auxin response factor 6 (*LOC112792550*), DELLA protein GAIP-B (*LOC112792521*). MQTL4.1 contains tomato spotted wilt virus, late leaf spot, and one palmitic acid QTL. F-box/LRR-repeat protein (*LOC112795816*), serine/threonine-protein kinase RUNKEL-like (*LOC112794155*), rust resistance kinase Lr10 (*LOC112795859*), receptor-like protein EIX2 (*LOC112795875*) are mostly related to the disease resistance and acyl carrier protein (*LOC112795780*) is related to the palmitic acid. MQTL5.3 contains disease- and oil-content QTLs. Some important genes related to these traits are proline-rich receptor-like protein kinase (*LOC112802144*), probable polyol transporter (*LOC112802148*), serine acetyltransferase 5-like (*LOC112803828*), serine/threonine-protein kinase phg2 (*LOC112803830*), zinc finger MYM-type protein 1-like (*LOC112803827*), protein FAR-RED IMPAIRED RESPONSE 1-like (*LOC112803829*).

In MQTL8.1, multiple disease-associated QTLs are clustered, and this region contains biotic stress responsive genes such as calcineurin B-like protein (*LOC112708056*), defensin-like protein 1 (*LOC112708031*), and pentatricopeptide repeat-containing protein At3g62890 (*LOC112708064*). On LG09, we generally found MQTLs (MQTL9.7 and MQTL9.3) associated with oil content. However, in MQTL9.7, we also found QTLs for morphological traits in addition to oil content. In MQTL9.3, the genes related to oil traits include fatty-acid desaturase FAD2 (*LOC112710390*), glycerol kinase (*LOC112710344*), linoleate 9S-lipoxygenase (*LOC112710463*), and serine carboxypeptidase (*LOC112710260*). Additionally, in MQTL9.7, genes related to plant growth and development were identified, including VAN3-binding protein (*LOC112711758*), DEAD-box ATP-dependent RNA helicase 7 (*LOC112711726*), ATP sulfurylase 1 (*LOC112711734*), and carboxylesterase 1 (*LOC112711707*). MQTL10.1 contains important genes responsive to biotic stress, including L-type lectin-domain containing receptor kinase VII.1-like (*LOC112715495*), polyadenylate-binding protein RBP47 (*LOC112715552*), serine carboxypeptidase-like 17 (*LOC112715547*), bZIP transcription factor 46 (*LOC112715565*), and zinc finger MYM-type protein 1-like (*LOC112717484*).

MQTL11.2 is the most important region, and it contains 14 QTLs. Squamosa promoter-binding-like protein 6 (*LOC112722612*), serine/threonine-protein phosphatase-7 long form homolog (*LOC112721816*), MADS-box protein (*LOC112722615*) are related to the pod weight. Zinc finger protein (*LOC112721803*), serine protease SPPA, F-box protein (*LOC112722634*) and isoaspartyl peptidase/L-asparaginase (*LOC112722621*) are related to the late leaf spot. Squamosa promoter-binding-like protein (*LOC112722612*), E3 ubiquitin ligase BIG BROTHER-related (*LOC112722633*), serine/threonine-protein phosphatase-7 long form homolog (*LOC112721816*) are related to the dry weight-related traits. Rhamnogalacturonan I rhamnosyl-transferase 1 (*LOC112722626*) and protein breaking of asymmetry in the stomatal lineage-like (*LOC112724059*) are related to the transcription rate and probable UDP-3-O-acylglucosamine N-acyltransferase (*LOC112722632*) is related to gadoleic acid content. These genes are also related to other

traits like haulm weight, 100 seed weight and pod constriction. QTLs in MQTL12.1 are related to the pod length and width and disease resistance (Tomato spotted wilt virus). The genes encode for gibberellin 20 oxidase (*LOC112728437*), receptor-like serine/threonine-protein kinase (*LOC112728351*), glyoxylate/hydroxy pyruvate reductase (*LOC112728355*), organelle RRM domain-containing protein (*LOC112728356*), disease resistance protein RML1B (*LOC112728387*), and TMV resistance protein N-like (*LOC112728444*) are mostly related to these traits. MQTL13.1 mostly contains transpiration and chlorophyll related QTLs, protein photosystem I assembly (*LOC112735782*), NDR1/HIN1-like protein (*LOC112737452*), pentatricopeptide repeat-containing protein (*LOC112737396*), protein LOW PSII accumulation 2 (*LOC112737399*), chloroplast envelope quinone oxidoreductase homolog (*LOC112737420*). On LG19, MQTL19.1 harbors QTLs for seed oil content and quality traits. It contains essential genes related to these oil-associated traits, including Acyl-CoA-binding domain-containing protein 6 (*LOC112776181*), fatty Acyl-CoA reductase FAR1 (*LOC112778876*), F-box proteins (*LOC112776494*), and cyclin-dependent kinase inhibitor CDKN1B (*LOC112776306*). MQTL14.3, MQTL18.8, MQTL19.1, MQTL19.4, and MQTL19.5 are found as a hotspot for most of the fatty acid's QTLs (Lignoceric acid, Stearic acid, Arachidic acid, Palmitic acid, Behenic acid, Gadoleic acid, Stearic acid, Arachidic acid, Oleic acid, Linoleic acid). This region contains multiple copies of oxysterol-binding protein-related protein (*LOC112741888*). Other trait related genes are TBC1 domain family member (*LOC112741897*), glucan endo-1,3-beta-glucosidase (*LOC112741948*), phosphatidylinositol/phosphatidylcholine transfer protein SFH11-like (*LOC112742554*).

4 Discussion

Developing high-yielding and climate-resilient crops is crucial for ensuring food and nutritional security. In recent years, various studies have been conducted on peanut to identify genomic regions associated with different important quantitative traits. Despite the identification of thousands of QTLs related to yield and other significant traits in various plant species, only a few have proven useful in genetic improvement programs due to their minor effects and environmental influences. Consequently, the primary objective of this MQTL analysis is to identify stable QTLs in plant genomes that can be effectively utilized in breeding programs through marker-assisted selection.

The distribution and precision of MQTLs across peanut chromosomes offer valuable insights into the genetic architecture of important agronomic traits. Our findings show that certain linkage groups, such as LG01, LG03 and LG05, harbor a higher concentration of MQTLs, indicating these chromosomes as critical regions for traits like disease resistance and seed weight. The presence of up to eight MQTLs on LG05, for instance, highlights its potential as a hotspot for genetic control of these key traits including aflatoxin resistance. The concentration of MQTLs in these linkage groups is comparable to previous studies in other legumes, such as soybean and common bean, where similar clustering of QTLs was observed in specific genomic regions. This suggests that certain chromosomes in legumes may play a more significant role in controlling a variety of traits, thereby making them prime targets for marker-assisted breeding.

A critical aspect of this study is the improved precision of MQTLs, as evidenced by the significant reduction in CIs compared to original QTLs. For example, MQTL3.3, with a CI as narrow as 0.07 cM, indicates a highly stable QTL region, a feature that significantly enhances the accuracy of breeding programs. This improvement in precision is also reflected in the overall reduction of CIs, with many linkage groups showing a fold change greater than four. This mirrors the outcomes of MQTL studies in other legumes, where the refinement of QTL regions has successfully reduced uncertainty around key genomic loci. The narrow CIs in MQTLs provide breeders with more reliable and specific targets, making it easier to incorporate these regions into marker-assisted selection (MAS) programs aimed at improving traits like yield, disease resistance, and seed quality.

In addition to trait-specific MQTLs, the presence of multi-trait MQTLs, such as MQTL5.4 harboring 15 QTLs associated with diverse traits like late leaf spot (LLS), oil content, and seed size, demonstrates the pleiotropic nature of certain genomic regions. This observation is consistent with previous reports in legumes, where specific MQTLs control multiple traits (Kumar et al., 2023). Such regions are particularly valuable in breeding programs as they enable the simultaneous improvement of multiple traits, reducing the need for separate selection processes for each characteristic. Another notable outcome of our study is the variation in marker density across different linkage groups. High-density regions, such as LG10 with 2.83 markers per cM, provide high-resolution genetic maps, which are crucial for fine-mapping QTLs and enhancing the precision of breeding strategies. In contrast, regions with lower marker density, like LG03 with 0.88 markers per cM, may require additional efforts to improve marker saturation for better trait mapping. The observed variability in marker density across linkage groups is consistent with findings in other legume species, where differences in marker distribution often reflect the complexity and evolutionary history of the genome.

Our study also highlights the complexity of nutrition-related traits, which had the highest number of QTLs (200), underscoring the importance of nutritional improvements in crop breeding. This aligns with the growing interest in enhancing nutrient content in legumes, a focus area shared with other legumes like soybean and chickpea, where nutritional traits like protein and oil content have been extensively studied. The identification of numerous QTLs for nutrient traits further emphasizes the potential for genetic improvement in this area, particularly as breeders aim to develop biofortified crops with enhanced nutritional profiles.

In our study, we identified important candidate genes related to the traits in the MQTL region. Total 48 genome-wide MQTLs were selected based on the availability of physical position of the flanking markers. Multiple disease resistance related genes like serine/threonine-protein kinase (*LOC112801194*), *BOI-related E3* ubiquitin-protein ligase (*LOC112801426*), autophagy-related protein (*LOC112802850*), auxin-responsive protein IAA30-like

(*LOC112718703*), WAT1-related protein (*LOC112802633*), protein YABBY 4 (*LOC112803178*), zinc finger MYM-type protein 1-like (*LOC112803827*), F-box/LRR-repeat protein (*LOC112795816*), rust resistance kinase Lr10 (*LOC112795859*), and receptor-like protein *EIX2* (*LOC112795875*) were identified from these MQTL regions. Sequencing-based bulk segregant analysis (seq-BSA) combined with nonsynonymous analysis identified the SNP variant in serine/threonine protein kinase gene which has significant role in the fusarium wilt and sterility mosaic disease resistance (Singh et al., 2016). In another study, the mutant population was screened to identify the role of *E3* ubiquitin-protein ligase gene for the disease resistance in rice. The mutant copy of this gene showed a higher expression against the disease infection (You et al., 2016). Transient silencing of the *WAT1* gene in tomato showed resistance to wilt pathogen (Hanika et al., 2021). The over deposition of the F-box proteins was recorded in the nucleus and cytoplasm of the wheat for the immunity activation signaling against the infection leaf rust pathogen (Wei et al., 2020). To protect the bread wheat against biotic stresses like pest and pathogens, wheat breeders have introduced over 200 genes into the cultivated gene pool from various sources. One of the disease resistance genes was tested using the transient expression for its contribution to the resistance against the stem rust. The overexpression of these receptor protein kinases showed resistance against the stem rust of wheat (Yu et al., 2023).

Similarly, transpiration related genes like gibberellin-regulated protein 14-like (*LOC112710127*), caffeoylshikimate esterase-like (*LOC112720721*), EPIDERMAL PATTERNING FACTOR-like protein (*LOC112754391*) and NAC domain-containing protein (LOC112767006) were identified. Gibberellin-regulated protein like DELLA plays a role in signaling the control of the transpiration rate for water use efficiency during the critical phases of plant development and stress management (Locascio et al., 2013). Caffeoylshikimate esterase plays a crucial role in lignin biosynthesis, which provides physical support and water protection by controlling the transpiration network pathway. Due to mutation in this gene and NAC domain, lignin deposition was hampered in one of the droughts related experiments in maize and Arabidopsis (Lu et al., 2013; Sun et al., 2018). Moreover, EPIDERMAL PATTERNING FACTOR-like protein affects the transpiration rate by regulating stomatal density and size. In rice, the overexpression of this gene negatively regulates and reduces the stomatal density, and after this scenario, the experimental lines of rice performed better in terms of their water use efficiency (Caine et al., 2019).

Gibberellin-regulated protein (*LOC112710127*), transcription factor *KUA1* (*LOC112802418*), and many small nuclear RNAs are related to the specific leaf area. GA controls the leaf size by regulating the cell division and increasing water absorption (Ritonga et al., 2023). *KUA* is an *MYB*-like transcription factor that activates leaf expansion and growth-related genes (Schmidt et al., 2016). The overexpression of this gene resulted in a larger cell size, while the mutant of this gene showed a reduction in the size of the cells in leaf tissue in Arabidopsis (Lu et al., 2014).

Iron and zinc content related genes are probable methyltransferase (*LOC112801274*), histone-lysine N-methyltransferase, H3 lysine-9 specific SUVH6 (*LOC112801997*), formyltetrahydrofolate deformylase 1 (*LOC112802537*), ferredoxin C 2 (*LOC112801048*), ion protease homolog 2 (*LOC112801369*), superoxide dismutase (*LOC112719733*) and NAC domain-containing protein (*LOC112767006*). Under iron deficiency, methyltransferase relaxes the chromatin structure for the sufficient expression of *nrf2* gene which regulates iron uptake (Su et al., 2022). Ferredoxin C and superoxidase dismutase is an iron binding protein in plants which also exoculate the transportation of iron (Sharma et al., 2023).

Oil content and fatty acid content related genes such as fatty-acid desaturase *FAD2* (*LOC112710390*), glycerol kinase (*LOC112710344*), linoleate 9S-lipoxygenase (*LOC112710463*), serine carboxypeptidase (*LOC112710260*), oxysterol-binding protein-related protein (*LOC112741888*), TBC1 domain family member (*LOC112741897*), glucan endo-1,3-beta-glucosidase (*LOC112741948*), phosphatidylinositol/phosphatidylcholine transfer protein SFH11-like (*LOC112742554*). Glycerol kinase provides the precursors to the lipid biosynthesis. The overexpression of this gene showed the increase in lipid production and showed the resistance to bacterial blight and blast diseases in rice (Xiao et al., 2022). Fatty-acid desaturase plays a crucial role lipid biosynthesis and for converting mono-unsaturated fatty acids into poly unsaturated fatty acids which is important for the normal development and function of plants (Czumaj and Śledziński, 2020). This enzyme also helps with other proteins in the determining the oil content and composition in the seeds. Linoleate 9S-lipoxygenase uses linoleic acid or linolenic acid as a substrate and gets involved in plant development and growth, stress tolerance, and senescence (Vellosillo et al., 2007). Phosphatidylinositol/phosphatidylcholine transfer protein and oxysterol-binding protein are also a lipid binding protein and play various roles in plant metabolism and stress response (Lete et al., 2020; Ye et al., 2022). The overexpression of the serine carboxypeptidase gene in the Arabidopsis negatively regulates and reduces the production of membrane lipid (Chen, 2020).

A few more genes related to plant growth and development were identified from different MQTL regions, including VAN3-binding protein (*LOC112711758*), DEAD-box ATP-dependent RNA helicase 7 (*LOC112711726*), ATP sulfurylase 1 (*LOC112711734*), and carboxylesterase 1 (*LOC112711707*). VAN3-binding protein activates proteins which regulate auxin transport mediated pathways and leads to continuous venation and root elongation (Naramoto and Kyozuka, 2018). DEAD-box ATP-dependent RNA helicase plays a role in plant growth, development, and found to be upregulated in abiotic stresses of wheat (Ru et al., 2021). ATP sulfurylase is important for Sulphur assimilation in plants. Sulphur is an essential macronutrient for the growth and development of plants (Anjum et al., 2015). Carboxylesterase also plays an important role in the growth, development, and stress tolerance in plants. The cis element of this gene was mostly found to be related to plant hormones like GA and IAA and the expression of this gene was mostly found in the root, leaf and stem of cotton (Rui et al., 2022).

In our study, while identifying candidate genes we have considered and highlighted the genes that are reported to play a key role in peanut as well as other key crops. For instance, genes encoding NAC domain-containing protein in the hotspot region

was reported to play key role in biotic, abiotic, nutritional, and physiological traits (Yuan et al., 2020; Li et al., 2021; Yuan et al., 2023). In addition, we identified key genes associated with aflatoxin resistance in MQTL5.2 (serine/threonine-protein kinase, BOI-related E3 ubiquitinprotein ligase), MQTL5.3, MQTL7.3, and MQTL13.1. Comparative proteomic studies indicated that serine/threonine-protein kinase was reported to play a role in aflatoxin resistance in maize (Chen et al., 2005; Brown et al., 2010). While global transcriptome profiling studies reported that BOI-related E3 ubiquitinprotein ligase plays a role on pre-harvest aflatoxin contamination in peanut. Similarly, for yield-related traits in MQTL3.1-MQTL3.4 (mitogen-activated protein kinase, auxin response factor), MQTL11.2 (MADS-box protein, squamosa promoter-binding protein), and MQTL14.1. *AhMPK3* exists in two copies in peanut genome and its structural organization revealed well-conserved nature of these signaling components across different species (Kumar et al., 2009). Further, auxin response factor *AhARF6* reported to play a role in peanut pod development (Li et al., 2021). In case of peanut, thirty-eight genes (*AhSPL1-AhSPL38*) were identified and *AhSPL* genes might be widely involved in peanut growth and development, as well as in response to environmental stresses (Sun et al., 2024). Several genes like fatty-acid desaturase FAD2, linoleate 9S-lipoxygenase, acyl-CoA-binding protein are reported to play role in the oil composition and some of these genes were edited and cloned (Peng et al., 2020; Neelakandan et al., 2022; Zhao et al., 2022). Genes related to oil composition within MQTL5.2 (fatty-acid desaturase FAD2, linoleate 9S-lipoxygenase), MQTL9.3, MQTL19.1 (acyl-CoA-binding protein, fatty acyl-CoA reductase FAR1), MQTL19.4, and MQTL19.5. Nutritional traits like iron and zinc content are linked to MQTL1.1 (probable methyltransferase, ferredoxin C), MQTL10.1, and MQTL12.1.

5 Conclusion

In conclusion, the outcomes of this study demonstrate the power of MQTL analysis in refining genomic regions associated with important traits. The reduction in QTL CIs, the identification of multi-trait MQTLs, and the concentration of MQTLs on key linkage groups provide valuable insights that can accelerate breeding efforts in peanut. These findings are consistent with MQTL studies in other legumes, suggesting a broader applicability of this approach across legume species. Future research should focus on further refining MQTL regions and increasing marker density in less-dense genomic regions to ensure comprehensive coverage of trait-associated loci, ultimately improving the effectiveness of MAS in crop breeding programs. Furthermore, we found that nutritional-related traits had the highest number of initial QTLs (200), reflecting their genetic complexity and the substantial research aimed at enhancing nutrient content in crops. Other categories, such as biotic traits (156 QTLs) and abiotic traits (57 QTLs), were also prominent, showing the broad scope of focus in breeding programs. When compared with previously reported MQTL studies in legumes, similar trends of concentration of MQTLs on key chromosomes and the improvement in QTL precision were observed. For example, studies on soybean and common bean also highlighted that specific chromosomes harbor denser MQTLs, particularly for traits like disease resistance and yield. The reduction in CIs in peanut mirrors findings in other legumes, indicating that MQTL analysis consistently enhances precision across legume species, providing breeders with robust, reliable genomic regions for trait improvement.

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

AS: Writing – original draft, Writing – review & editing, Formal analysis, Investigation. SR: Formal analysis, Investigation, Writing – original draft, Writing – review & editing. YN: Investigation, Writing – original draft, Writing – review & editing. AJ: Writing – original draft, Writing – review & editing, Formal analysis. MP: Writing – original draft, Writing – review & editing, Resources. RV: Resources, Writing – original draft, Writing – review & editing. MT: Conceptualization, Investigation, Supervision, Writing – original draft, Writing – review & editing, Funding acquisition. SP: Conceptualization, Funding acquisition, Supervision, Writing – original draft, Writing – review & editing, Resources.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. Evans-Allen project 336129-GEOX-5221 awarded to SP and USDA ARS Dawson peanut funding-Agreement Number/FAIN: 58-6044-1-003-Project No-6044-21000-005-006S-FVSU project-335149. Thanks for funding support received by ICRISAT in conducting different genetic mapping studies by Bill and Melinda Gates Foundation (BMGF), USA and Indian Council of Agricultural Research (ICAR), India.

Conflict of interest

The reviewer AK declared a past co-authorship with the authors YN, MT and SMP to the handling editor.

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

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

Publisher's note

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.2025.1539641/full#supplementary-material

References

Agropedia (2023). Groundnut: its importance. Available online at: http://agropedia.iitk.ac.in/content/groundnut-its-importance (Accessed November 2024).

Ahmad, S., Nawade, B., Sangh, C., Mishra, G. P., Bosamia, T. C., Kumar, N., et al. (2020). Identification of novel QTLs for late leaf spot resistance and validation of a major rust QTL in peanut (*Arachis hypogaea* L.). *Biotech* 10, 458. doi: 10.1007/s13205-020-02446-4

Anjum, N. A., Gill, R., Kaushik, M., Hasanuzzaman, M., Pereira, E., Ahmad, I., et al. (2015). ATP-sulfurylase, sulfur-compounds, and plant stress tolerance. *Front. Plant Sci.* 6. doi: 10.3389/fpls.2015.00210

APEDA (2024). Available at: https://apeda.gov.in/.

Arcade, A., Labourdette, A., Falque, M., Mangin, B., Chardon, F., Charcosset, A., et al. (2004). BioMercator: integrating genetic maps and QTL towards discovery of candidate genes. *Bioinform. (Oxford England)* 20 (14), 2324–2326. doi: 10.1093/bioinformatics/bth230

Bertioli, L., Soraya, C. M., Moretzsohn, M. C., Roberts, P. A., Ballén-Taborda, C., Borba, T., et al. (2016). Genetic mapping of resistance to meloidogyne arenaria in arachis stenosperma: a new source of nematode resistance for peanut. *G3: Genes Genomes Genet.* 6 (2), 377–390. doi: 10.1534/g3.115.023044

Brown, R. L., Chen, Z. Y., Warburton, M., Luo, M., Menkir, A., Fakhoury, A., et al. (2010). Discovery and characterization of proteins associated with aflatoxin-resistance: evaluating their potential as breeding markers. *Toxins (Basel).* 2, 919–933. doi: 10.3390/toxins2040919

Caine, R. S., Yin, X., Sloan, J., Harrison, E. L., Mohammed, U., Fulton, T., et al. (2019). Rice with reduced stomatal density conserves water and has improved drought tolerance under future climate conditions. *New Phytol.* 221, 371–384. doi: 10.1111/nph.15344

Chen, J. (2020). The serine carboxypeptidase-like gene SCPL41 negatively regulates membrane lipid metabolism in *Arabidopsis thaliana*. *Plants* 9, 696. doi: 10.3390/plants9060696

Chen, Z. Y., Brown, R. L., Menkir, A., Damann, K. E., and Cleveland, T. E. (2005). Proteome analysis of near isogenic maize lines differing in the level of resistance against Aspergillus flavus infection/aflatoxin production. *Phytopathology* 95, S19. doi: 10.1094/PHYTO-96-0087

Czumaj, A., and Śledziński, T. (2020). Biological role of unsaturated fatty acid desaturases in health and disease. *Nutrients* 12, 356. doi: 10.3390/nu12020356

Darvasi, A., and Soller, M. A. (1997). Simple method to calculate resolving power and confidence interval of QTL map location. *Behav Genet.* 27 (2), 125–132. doi: 10.1023/a:1025685324830

Dodia, S. M., Joshi, B., Gangurde, S. S., Thirumalaisamy, P. P., Mishra, G. P., Narandrakumar, D., et al. (2018). Genotyping-by-sequencing based genetic mapping reveals large number of epistatic interactions for stem rot resistance in groundnut. *Theor. Appl. Genet.* 132 (4), 1001–1016. doi: 10.1007/s00122-018-3255-7

Endelman, J. B., and Plomion, C. (2014). LPmerge: an R package for merging genetic maps by linear programming. *Bioinform. (Oxford England).* 30 (11), 1623–1624. doi: 10.1093/bioinformatics/btu091

FAOSTAT (2021). Available at: https://www.fao.org/faostat/en/#home.

Gangurde, S. S., Xavier, A., Naik, Y. D., Jha, U. C., Rangari, S. K., Kumar, R., et al. (2022). Two decades of association mapping: Insights on disease resistance in major crops. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.1064059

Guo, S., Zhang, Y., Yuan, F., Gao, Y., Gu, L., Wong, I., et al. (2006). Regulation of replication protein a functions in DNA mismatch repair by phosphorylation. *J. Biol. Chem.* 281 (31), 21607–21616. doi: 10.1074/jbc.M603504200

Gupta, S., Rangari, S. K., Sahu, A., Naik, Y. D., Satayavathi, C. T., Punnuri, S., et al. (2024). Meta-QTL analysis reveals the important genomic regions for biotic stresses, nutritional quality and yield-related traits in pearl millet. *CABI Agric. Biosci.* 5, 36. doi: 10.1186/s43170-024-00230-5

Hanika, K., Schipper, D., Chinnappa, S., Oortwijn, M., Schouten, H. J., Thomma, P. H. J., et al. (2021). Impairment of tomato WAT1 enhances resistance to vascular wilt fungi despite severe growth defects. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.721674

Kumar, R., Sharma, V. K., Rangari, S. K., Jha, U. C., Sahu, A., Paul, P. J., et al. (2023). High confidence QTLs and key genes identified using Meta-QTL analysis for enhancing heat tolerance in chickpea (*Cicer arietinum* L.). *Front. Plant Sci.* 14. doi: 10.3389/fpls.2023.1274759

Kumar, K. R. R., Srinivasan, T., and Kirti, P. B. (2009). A mitogen-activated protein kinase gene, AhMPK3 of peanut: molecular cloning, genomic organization, and heterologous expression conferring resistance against Spodoptera litura in tobacco. *Mol. Genet. Genomics* 282, 65–81. doi: 10.1007/s00438-009-0446-6

Lete, M. G., Tripathi, A., Chandran, V., Bankaitis, V. A., and McDermott, M. I. (2020). Lipid transfer proteins and instructive regulation of lipid kinase activities: Implications for inositol lipid signaling and disease. *Adv. Biol. Regul.* 78, 100740. doi: 10.1016/j.jbior.2020.100740

Liao, B. S. (2003). *The groundnut* (Wuhan: Hubei Press for Science and Technology).

Locascio, A., Blázquez, M. A., and Alabadí, D. (2013). Genomic analysis of DELLA protein activity. *Plant Cell Physiol.* 54, 1229–1237. doi: 10.1093/pcp/pct082

Lu, S., Li, Q., Wei, H., Chang, M. J., Tunlaya-Anukit, S., and Kim, H. (2013). Ptr-miR397a is a negative regulator of laccase genes affecting lignin content in Populus trichocarpa. *Proc. Natl. Acad. Sci. U.S.A.* 110, 10848–10853. doi: 10.1073/pnas.1308936110

Lu, Q., Liu, H., Hong, Y., and Liang, X. (2018). Consensus map integration and QTL meta-analysis narrowed a locus for yield traits to 0.7 cM and refined a region for late leaf spot resistance traits to 0.38 cM on linkage group A05 in peanut (Arachis hypogaea L.). *BMC Genomics* 19, 887. doi: 10.1186/s12864-018-5288-3

Lu, D., Wang, T., Persson, S., Mueller-Roeber, B., and Schippers, H. M. (2014). Transcriptional control of ROS homeostasis by KUODA1 regulates cell expansion during leaf development. *Nat. Commun.* 5, 3767. doi: 10.1038/ncomms4767

Miceli, A., Moncada, A., Sabatino, L., and Vetrano, F. (2019). Effect of gibberellic acid on growth, yield, and quality of leaf lettuce and rocket grown in a floating system. *Agronomy* 9, 382. doi: 10.3390/agronomy9070382

Naramoto, S., and Kyozuka, J. (2018). ARF GTPase machinery at the plasma membrane regulates auxin transport-mediated plant growth. *Plant Biotechnol.* 35, 155–159. doi: 10.5511/plantbiotechnology.18.0312a

National Peanut Board (2023). Peanuts are a crop for now. Available online at: https://www.nationalpeanutboard.org/news/peanuts-are-crop-now (Accessed November 2024).

Neelakandan, A. K., Wright, D. A., Traore, S. M., Chen, X., Spalding, M. H., and He, G. (2022). CRISPR/Cas9 based site-specific modification of FAD2 cis-regulatory motifs in peanut (Arachis hypogaea L). *Front. Genet.* 13, 849961. doi: 10.3389/fgene.2022.849961

Olivier, S., Charcosset, A., and Joets, J. (2012). BioMercator V3: An upgrade of genetic map compilation and quantitative trait loci meta-analysis algorithms. *Bioinform.* 28 (15), 2082–2083. doi: 10.1093/bioinformatics/bts313

Pandey, M. K., Pandey, A. K., Kumar, R., Nwosu, C. V., Guo, B., Wright, G. C., et al. (2020). Translational genomics for achieving higher genetic gains in groundnut. *Theor. Appl. Genet.* 133, 1679–1702. doi: 10.1007/s00122-020-03592-2

Pandey, M. K., Wang, H., Khera, P., Vishwakarma, M. K., Kale, S. M., Culbreath, A. K., et al. (2017). Genetic dissection of novel QTLs for resistance to leaf spots and tomato spotted wilt virus in peanut (Arachis hypogaea L.). *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.00025

Pandey, M. K., Wang, M. L., Qiao, L., and Qin, H. (2014). Identification of QTLs associated with oil content and mapping FAD2 genes and their relative contribution to oil quality in peanut (Arachis hypogaea L.). *BMC Genet.* 15, 133. doi: 10.1186/s12863-014-0133-4

PeanutBase (2023). PeanutBase: genome resources for peanut researchers. Available online at: https://www.peanutbase.org/ (Accessed November 2024).

Peng, Z., Ruan, J., Tian, H., Shan, L., Meng, J., Guo, F., et al. (2020). The family of peanut fatty acid desaturase genes and a functional analysis of four ω-3 AhFAD3 members. *Plant Mol. Biol. Rep.* 38, 209–221. doi: 10.1007/s11105-019-01191-0

Ritonga, F. N., Zhou, D., Zhang, Y., Song, R., Li, C., Li, J., et al. (2023). The roles of gibberellins in regulating leaf development. *Plants* 12, 1245. doi: 10.3390/plants12061243

Ru, J., Hou, Z., Zheng, L., Zhao, Q., Wang, F., Chen, J., et al. (2021). Genome-wide analysis of DEAD-box RNA helicase family in wheat (Triticum aestivum) and functional identification of TaDEAD-box57 in abiotic stress responses. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.797276

Rui, C., Peng, F., Fan, Y., Zhang, Y., Zhang, Z., Xu, N., et al. (2022). Genome-wide expression analysis of carboxylesterase (CXE) gene family implies GBCXE49 functional responding to alkaline stress in cotton. *BMC Plant Biol.* 22, 194. doi: 10.1186/s12870-022-03579-9

Schmidt, R., Kunkowska, A. B., and Schippers, H. M. (2016). Role of reactive oxygen species during cell expansion in leaves. *Plant Physiol.* 172, 2098–2106. doi: 10.1104/pp.16.00426

Sharma, A., Sharma, D., and Verma, S. K. (2023). A systematic in silico report on iron and zinc proteome of *Zea mays*. *Front. Plant Sci.* 14. doi: 10.3389/fpls.2023.1166720

Shasidhar, Y., Vishwakarma, M. K., Pandey, M. K., Janila, P., Variath, M. T., Manohar, S. S., et al. (2017). Molecular mapping of oil content and fatty acids using dense genetic maps in groundnut (*Arachis hypogaea* L.). *Front. Plant Sci.* 8, 258666. doi: 10.3389/fpls.2017.00794

Singh, V. K., Khan, A. W., Saxena, R. K., Kumar., V., Kale, S. M., Sinha, P., et al. (2016). Next-generation sequencing for identification of candidate genes for Fusarium wilt and sterility mosaic disease in pigeonpea (*Cajanus cajan*). *Plant Biotechnol. J.* 14, 1183–1194. doi: 10.1111/pbi.12470

Su, J., Yao, Z., Wu, Y., Lee, J., and Jeong, J. (2022). Minireview: Chromatin-based regulation of iron homeostasis in plants. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.959840

Sujay, V., Gowda, M. V. C., Pandey, M. K., Bhat, R. S., Khedikar, Y. P., Nadaf, H. L., et al. (2012). Quantitative trait locus analysis and construction of consensus genetic map for foliar disease resistance based on two recombinant inbred line populations in cultivated groundnut (*Arachis hypogaea* L.). *Mol. Breed.* 30, 773–788. doi: 10.1007/s11032-011-9661-z

Sun, Q., Liu, X., Yang, J., Liu, W., Du, Q., Wang, H., et al. (2018). MicroRNA528 affects lodging resistance of maize by regulating lignin biosynthesis under nitrogen-luxury conditions. *Mol. Plant* 11, 806–814. doi: 10.1016/j.molp.2018.03.013

Sun, X., Zhang, L., Xu, W., Zheng, J., Yan, M., Zhao, M., et al. (2024). A comprehensive analysis of the peanut SQUAMOSA promoter binding protein-like gene family and how *ahSPL5* enhances salt tolerance in transgenic *arabidopsis*. *Plants* 13, 1057. doi: 10.3390/plants13081057

USDA (2024). Available at: https://ipad.fas.usda.gov/cropexplorer/cropview/commodityView.aspx?cropid=2221000.

Variath, M. T., and Janila, P. (2017). "Economic and academic importance of peanut," in *The peanut genome. compendium of plant genomes*. Eds. R. Varshney, M. Pandey and N. Puppala (Cham: Springer). doi: 10.1007/978-3-319-63935-2_2

Vellosillo, T., Miguel, A., Jorge, V., Liam, D., Mats, H., and Carmen, C. (2007). Oxylipins produced by the 9-lipoxygenase pathway in *Arabidopsis* regulate lateral root development and defense responses through a specific signaling cascade. *Plant Cell* 19 (3), 831–846. doi: 10.1105/tpc.106.046052

Venske, E., Dos Santos, R. S., Farias, D. D. R., Rother, V., da Maia, L. C., Pegoraro, C., et al. (2019). Meta-analysis of the QTLome of fusarium head blight resistance in bread wheat: Refining the current puzzle. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.00727

Veyrieras, J. B., Gofinet, B., and Charcosset, A. (2007). MetaQTL: A package of new computational methods for the meta-analysis of QTL mapping experiments. *BMC Bioinforma.* 8, 49.

Vishwakarma, M. K., Nayak, S. N., Guo, B., Wan, N., Liao, B., Varshney, R. K., et al. (2017). "Classical and molecular approaches for mapping of genes and quantitative trait loci in peanut," in *The peanut genome*. Eds. R. K. Varshney, M. K. Pandey and N. Puppala (Cham, Switzerland: Springer International Publishing), 93–116. doi: 10.1007/978-3-319-63935-2_7

Wang, M. L., Khera, P., Pandey, M. K., Wang, H., Qiao, L., Feng, S., et al. (2015). Genetic mapping of QTLs controlling fatty acids provided insights into the genetic control of fatty acid synthesis pathway in peanut (*Arachis hypogaea* L.). *PloS One* 10, e0119454. doi: 10.1371/journal.pone.0119454

Wei, J., Cui, L. P., Zhang, N., Du, D. D., and Yang, W. X. (2020). Puccinia triticina pathotypes THTT and THTS display complex transcript profiles on wheat cultivar thatcher. *BMC Genet.* 21, 48. doi: 10.21203/rs.2.12168/v4

Xiao, X., Wang, R., Khaskhali, S., Gao, Z., Guo, W., Wang, H., et al. (2022). A novel glycerol kinase gene OsNHO1 regulates resistance to bacterial blight and blast diseases in rice. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.800625

Ye, H., Gao, J., Liang, Z., Lin, Y., Yu, Q., Huang, S., et al. (2022). Arabidopsis ORP2A mediates ER–autophagosomal membrane contact sites and regulates PI3P in plant autophagy. *Proc. Natl. Acad. Sci. U.S.A.* 119, e2205314119. doi: 10.1073/pnas.2205314119

You, Q., Zhai, K., Yang, D., Yang, W., Wu, J., Liu, J., et al. (2016). An E3 ubiquitin ligase-BAG protein module controls plant innate immunity and broad-spectrum disease resistance. *Cell Host Microbe* 20, 758–769. doi: 10.1016/j.chom.2016.10.023

Yu, B., Huai, D., Huang, L., Kang, Y., Ren, X., Chen, Y., et al. (2019). Identification of genomic regions and diagnostic markers for resistance to aflatoxin contamination in peanut (*Arachis hypogaea* L.). *BMC Genet.* 20, 32. doi: 10.1186/s12863-019-0734-z

Yu, G., Matny, O., Gourdoupis, S., Rayapuram, N., Aljedaani, F. R., Wang, Y. L., et al. (2023). The wheat stem rust resistance gene *Sr43* encodes an unusual protein kinase. *Nat. Genet.* 55, 921–926. doi: 10.1038/s41588-023-01402-1

Yuan, C., Li, C., Lu, X., Zhao, X., Yan, C., Wang, J., et al. (2020). Comprehensive genomic characterization of NAC transcription factor family and their response to salt and drought stress in peanut. *BMC Plant Biol.* 20, 1–21. doi: 10.1186/s12870-020-02678-9

Yuan, X., Wang, Y., Ji, P., Wu, P., Sheffield, J., and Otkin, J. A. (2023). A global transition to flash droughts under climate change. *Sci.* doi: abn6301

Zhao, S., Sun, J., Sun, J., Zhang, X., Zhao, C., Pan, J., et al. (2022). Insights into the novel FAD2 gene regulating oleic acid accumulation in peanut seeds with different maturity. *Genes* 13, 2076. doi: 10.3390/genes13112076

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 03 June 2025
DOI 10.3389/fpls.2025.1595870

Check for updates

OPEN ACCESS

EDITED BY
Somashekhar M Punnuri,
Fort Valley State University, United States

REVIEWED BY
Digvijay Verma,
Babasaheb Bhimrao Ambedkar University, India
Ashish Kumar,
Jawaharlal Nehru Agricultural University, India
Jian Xiao,
Hunan University, China

*CORRESPONDENCE
Pan Luo
luopan@hubu.edu.cn
Jun Yu
yujun80324@163.com
Chunlei Yang
ycl193737@163.com

†These authors have contributed equally to this work

RECEIVED 18 March 2025
ACCEPTED 06 May 2025
PUBLISHED 03 June 2025

CITATION
Liu X, Qiao Y, Wu X, Chen X, Yang F, Li H, Li C, Yang Y, Yang C, Yu J and Luo P (2025) Comprehensive metabolomic and microbial analysis of tobacco rhizosphere soil responses to crop rotation and fertilization. *Front. Plant Sci.* 16:1595870. doi: 10.3389/fpls.2025.1595870

COPYRIGHT
© 2025 Liu, Qiao, Wu, Chen, Yang, Li, Li, Yang, Yang, Yu and Luo. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Comprehensive metabolomic and microbial analysis of tobacco rhizosphere soil responses to crop rotation and fertilization

Xinyu Liu[1†], Ying Qiao[1†], Xianzhi Wu[2†], Xuanxuan Chen[3†], Fan Yang[1†], Hao Li[3], Chuanzong Li[3], Yong Yang[1], Chunlei Yang[3*], Jun Yu[3*] and Pan Luo[1*]

[1]State Key Laboratory of Biocatalysis and Enzyme Engineering, College of Life Sciences, Hubei University, Wuhan, China, [2]Bijie Academy of Agricultural Sciences, Bijie, China, [3]Tobacco Research Institute of Hubei Province, Wuhan, China

Tobacco (*Nicotiana tabacum* L.) is a crucial Solanaceae crop globally, but its continuous cultivation can lead to soil degradation. Crop rotation offers numerous benefits, including enhanced soil fertility, improved microbial communities, and pest control. However, how different tobacco planting systems specifically reshape rhizosphere metabolite profiles and regulate microbial diversity remains unclear. Here, we analyzed soil samples from four tobacco cropping systems using non-targeted metabolomics, 16S rRNA and ITS sequencing. The results revealed distinct changes in soil metabolite profiles and microbial communities under different treatments. We identified significant alterations in lipid metabolism, amino acid biosynthesis, and secondary metabolite pathways, which influence soil microbial populations and tobacco plant health. Lipid metabolites, including fatty acids and eicosanoids, were particularly notable for their roles in microbial signaling and plant defense. Furthermore, microbial gene abundance analysis indicated that different treatments fostered unique microbial populations, including increased arbuscular mycorrhizal fungi and saprotrophic fungi, which support nutrient cycling and plant growth. These findings highlight the critical interplay between soil metabolites, microbial diversity, and plant productivity, offering insights into optimizing tobacco cropping systems for improved soil health and sustainable agricultural practices.

KEYWORDS

tobacco, crop rotation, rhizosphere metabolites, microbial diversity, soil fertility

1 Introduction

Tobacco (*Nicotiana tabacum* L.), a member of the Solanaceae family, is one of the most important crops worldwide due to its economic value (Chen et al., 2022). While continuous tobacco cropping offers advantages such as labor savings, reduced working time, and lower production costs, it can also cause long-term soil degradation (Xiong et al., 2023). To mitigate these issues, crop rotation is a common practice that alters the dynamics of tobacco rhizosphere metabolites, impacting soil fertility (Zhou et al., 2023a). Rotation improves soil organic matter content by promoting the decomposition of previous crop residues and enhancing the soil microbial community (Tiemann et al., 2015). It also optimizes soil physical and chemical properties, such as pH and structure, and increases nutrient availability, improving overall nutrient cycling (Behnke et al., 2021). Furthermore, crop rotation helps control soil-borne diseases and pests, ensuring a healthier growth environment for tobacco plants (Lu et al., 2022).

Strip intercropping, a synergistic agroecological system, stands as one of the most efficacious strategies for sustainable yield intensification. *Vulpia myuros* (L.) C.C. Gmelin is a commonly utilized plant for intercropping with crops due to its ecological benefits. Intercropping tobacco with *V. myuros* offers several benefits, including improved soil structure, suppression of weed growth, better regulation of soil nutrients, and reduced pest and disease incidence (Zhang et al., 2020). Similarly, rotating tobacco with crops like rape (*Brassica napus*) enhances soil health, provides pest control, and boosts economic returns (Liu et al., 2024a). However, excessive use of compound fertilizers in tobacco farming can lead to nutrient accumulation, particularly nitrogen and phosphorus, causing soil pollution over time (Ahmed et al., 2017).

The rotation of tobacco with *V. myuros* and rape has distinct effects on soil health, pest and disease control, and economic benefits (Fang et al., 2016; Zhang et al., 2020). These differences reflect variations in soil fertility, pest management strategies, cost-efficiency, and ecological sustainability (Liu et al., 2024b). In tobacco cultivation, two commonly used types of fertilizers are compound fertilizers and fermented cake fertilizers (Jiang et al., 2022). Compound fertilizer refers to a chemical fertilizer containing at least two of the three primary nutrients-nitrogen (N), phosphorus (P), and potassium (K) (Chen et al., 2024a). It is manufactured through chemical synthesis or physical blending, characterized by high nutrient density, minimal secondary components, and stable physicochemical properties. Compound fertilizers, with their balanced nutrient composition, provide rapid and stable fertilization, but their long-term use can degrade soil structure and, if misapplied, may cause seedling damage (Jiang et al., 2022). Fermented cake fertilizer refers to an organic fertilizer produced through microbial fermentation and decomposition of oilseed residues after oil extraction (such as Brassica napusseed cake, soybean cake, and peanut cake). This material serves dual functions by providing organic matter and essential nutrients (e.g., nitrogen, phosphorus, and potassium), thereby enhancing soil fertility and plant growth. Fermented cake fertilizers, on the other hand, improve soil fertility and structure and promote microbial activity. Though their effects are slower, they enhance the aroma and quality of tobacco (Feng et al., 2024). However, their use requires careful management due to their lower nutrient content and cumbersome application process (Feng et al., 2024).

Soil rhizosphere metabolites and microbial communities play a crucial role in linking soil health and crop productivity (Bhattacharyya and Jha, 2012). The impact of different planting methods on tobacco growth is largely mediated by changes in soil metabolites and microbial populations (Sun et al., 2023a). Understanding the relationship between soil metabolites, microbial diversity, and the functional capabilities of microbial communities could offer valuable insights into the advantages of various planting systems for tobacco cultivation (Holden, 2019).

This study investigated four tobacco cropping systems through integrated rhizosphere analysis. Soil samples were systematically collected from distinct cultivation regimes, employing non-targeted metabolomics to characterize metabolites alongside 16S (16S rRNA) and ITS (Internal Transcribed Spacer) amplicon sequencing for deciphering microbial diversity and identifying keystone taxa. Fertilization and intercropping systems may differentially regulate soil microbiome functions. Fermented cake fertilizer enhances bacterial/fungal richness through high active carbon input. V.myuros intercropping optimizes nitrogen cycling and carbon transformation through plant microbe interactions, while rapeseed rotation promotes phosphorus activation and pathogen inhibition. Microbial function differences are driven by lignin/cellulose ratio. The rationale for comparing key metabolites and microbial interactions lies in their dual role as indicators and drivers of soil health. By elucidating how agricultural practices remodel rhizosphere metabolic landscapes and microbial consortia, this study systematically deciphers the mechanisms underlying soil-microbe-plant interactions, providing a new perspective to optimize tobacco cropping systems for enhanced soil functionality and ecological sustainability.

2 Materials and methods

2.1 Plant materials, growing conditions and treatments

The location of the test is in Xijiadian County (111.18354°E, 32.748336°N), Danjiangkou City, Hubei Province, China. The soil type is classified as yellow-brown earth. Tobacco cultivar of the Yunyan 87 (*Nicotiana tabacum* cv. Yunyan 87) was provided by Wuhan Tobacco Research Institute, Hubei Province. Rhizosphere soil samples were systematically collected using a standardized five-point sampling protocol. For each experimental group, a minimum of six individual plants were selected to ensure representative sampling. Following collection, all samples were promptly transferred to the laboratory under cryogenic conditions and preserved at −80°C to maintain biochemical integrity prior to downstream analyses.

Control treatment (CK): In the CK, ridge cultivation is performed with conventional fertilization to meet the nutrient

requirements of tobacco cultivation. The conventional fertilization treatment was a compound fertilizer with nitrogen of 16 kg (35 kg of urea, 46% of nitrogen), P_2O_5 of 5.4 kg (30 kg calcium superphosphate, 18% of phosphorus) and K_2O of 5 kg (10 kg potassium sulfate, 50% of K_2O) and applied as 75 kg ha^{-1} during the growing season. Additionally, the compound fertilizer was applied at a rate of 1.5 kg per row, with rows spaced 1.2 m apart. Fertilizers were uniformly broadcasted and incorporated into the top 20 cm soil layer during ridge preparation, consistent with regional agricultural practices. Plot - based positioning is adopted, and monitoring is continuously carried out over a five - year period (Zhu et al., 2023).

Treatment 1 (T1, rotation + intercropping group): During late September 2021, *V. myuros* is sown through broadcast seeding at a rate of 300 g of seeds per plot. Subsequently, in late March 2022, the land is plowed and ridges are formed during the fertilization process, with 1.0 kg of compound fertilizer per row. In June 2022, the desiccated above - ground components of *V. myuros* among the ridges are relocated onto the ridges through hoeing for soil - heaping. Each year, a rotational and cyclic planting regime between tobacco and *V. myuros* rows is enforced (Rotation + Intercropping group) (Zhang et al., 2020).

Treatment 2 (T2, rotation group): Conventional fertilization - based ridge cultivation is carried out, with 1.0 kg of compound fertilizer per row. At the end of September, green manure (300 g of rapeseed per plot) is broadcast - sown, and at the end of March, the green manure is turned under for soil improvement (Fang et al., 2016).

T3 (Treatment 3, cake fertilizer group): Ridge cultivation with conventional fertilization is practiced, with 1.0 kg of compound fertilizer and 1 kg of fermented cake fertilizer per row respectively.

2.2 Quantification of functional gene abundances

The PCR products were purified using magnetic bead purification. Samples were mixed in equidensity ratios based on the concentration of PCR products. After thorough mixing, the PCR products were detected and target bands were recovered (Fatima et al., 2014). Sequencing libraries were generated and indexes were added. The library was checked with Qubit and real-time PCR for quantification and bioanalyzer for size distribution detection. Quantified libraries were pooled and sequenced on Illumina platforms, according to effective library concentration and data amount required (Bokulich et al., 2013).

We employed a multi-region primer amplification strategy targeting bacterial 16S (V3/V4/V4-V5), archaeal 16S (ArcV4), and fungal ITS (ITS1/ITS2) regions using barcoded primers (e.g., 515F-806R for 16SV4). PCR reactions utilized Phusion High-Fidelity Master Mix with thermal cycling: 98°C/1 min initial denaturation, 30 cycles (98°C/10 s, 50°C/30 s, 72°C/30 s), and 72°C/5 min final extension. PCR products were purified via magnetic beads, pooled by concentration, and validated before library preparation. Libraries were quantified using Qubit and qPCR, then sequenced on Illumina platforms (Bokulich et al., 2013). Bioinformatics analysis included: 1) demultiplexing reads by barcode/primer trimming; 2) FLASH-based paired-end assembly into Raw Tags; 3) fastp-filtered Clean Tags; 4) chimera removal via SILVA (16S) and UNITE (ITS) database alignment using vsearch, generating Effective Tags for downstream analysis (Magoč and Salzberg, 2011).

We processed Effective Tags using the QIIME2 software: denoising was performed via DADA2 or deblur modules to generate Amplicon Sequence Variants (ASVs, default: DADA2). Species annotation was conducted in QIIME2, with Silva Database for 16S rRNA genes, UNITE Database for ITS regions, and the Micro_NT sub-library (extracted from NT database for archaea, fungi, viruses, and bacteria) for non-standard genomic regions. The primer sequences used for each region were listed in Supplementary Table S1.

2.3 Statistical analysis

Paired-end reads were merged using FLASH (v1.2.11, http://ccb.jhu.edu/software/FLASH/) (Magoč and Salzberg, 2011), a very fast and accurate analysis tool, which was designed to merge paired end reads when at least some of the reads overlap theread generated from the opposite end of the same DNA fragment, and the splicing sequences were called raw tags (Magoč and Salzberg, 2011). Quality filtering on the raw tags was performed using the fastp (v0.23.1) software to obtain high-quality Clean Tags (Bokulich et al., 2012). The tags were compared with the reference database [Silva database (16S), https://www.arb-silva.de/; Unite Database (ITS), https://unite.ut.ee/] to detect chimera sequences, and the effective tags were obtained by removing the chimera sequences with the vsearch package (v2.16.0) (https://github.com/torognes/vsearch) (Edgar et al., 2011). For the Effective Tags obtained previously, denoising was performed with DADA2 or deblur module in the QIIME2 software to obtain initial ASVs (Amplicon Sequence Variants) (default: DADA2). Species annotation was performed using QIIME2 software. For 16S, the annotation database is Silva Database, while for ITS, it is Unite Database. For the un- regular region, the default database is Micro_NT (a sub library obtained by extracting archaea, fungi, viruses, and bacteria from the NT) (Yan et al., 2019).

A series of statistical analyses which include Anosim, Adonis, Multi-response permutation procedure (MRPP), Simper, T-test, MetagenomeSeq and LEfSe, was performed to reveal the community structure differentiation. Anosim, Adonis and MRPP analyses are non-parametric tests that analyze the differences between high-dimensional data groups. They can test whether the differences between groups are significantly greater than the differences within the group, which can determine whether the grouping is meaningful. All of them were performed with vegan and ggplot2 package within R. Simper can reveal the contribution of each species to the differentiation between groups. Top 10 species were selected and presented on the graph. It was performed in R with Vegan package and ggplot2 package. MetagenomeSeq can

showcase the species that display significant differences between groups. It was performed in R with metagenomeSeq package. LEfSe is widely used to discover biomarkers and it can reveal metagenomic characteristics. To achieve this, an exclusive package named lefse was utilized (Bachy et al., 2013).

To explore the symbiotic relationship between species and to reveal the environmental factor influence on the community structures, 2D and 3D network diagrams were drawn for visualization. Further analyses such as spearman correlation test, canonical correspondence analysis (CCA)/redundancy analysis (RDA) and dbRDA can be used to reflect the correlation between environmental factors and species abundance. All of these diagrams and analysis were completed in R.

Top 10 taxa of each sample at each taxonomic rank (Phylum, Class, Order, Family, Genus, Species) were selected to plot the relative abundance distribution histogram using the SVG function in Perl. We chose barplots of Phylum, Genus and Species levels for analysis.

In order to analyze the diversity, richness and uniformity of the communities in the sample, alpha diversity was calculated from 7 indices in QIIME2, including Observed_otus, Chao1, Shannon, Simpson, Dominance, Good's coverage and Pielou_e. We chose Shannon and Chao1 for analysis.In order to evaluate the richness of microbial community and determine the appropriate sample size, the species accumulation boxplot can be used for visualize, which was performed with the vegan package in R software (Bokulich et al., 2018).

Simper can reveal the contribution of each species to the differentiation between groups. Top 10 species were selected and presented on the graph. It was performed in R with Vegan package and ggplot2 package.

2.4 Untargeted metabolomics

Metabolites were extracted from tissue (100 mg homogenized in liquid nitrogen), liquid (100 µL), cell/bacteria, or culture medium samples using prechilled 80% methanol, followed by ice incubation, centrifugation (15,000 g, 4°C), and dilution to 53% methanol prior to LC-MS/MS injection. UHPLC-MS/MS analysis utilized a Vanquish UHPLC system coupled with Orbitrap Q Exactive™ HF/HF-X mass spectrometer (Thermo Fisher) equipped with a Hypersil Gold column (100×2.1 mm, 1.9 µm) under a 12-min gradient (0.2 mL/min; mobile phase: 0.1% FA in water [A] and methanol [B]). Data processing via Compound Discoverer 3.3 included peak alignment, normalization (total spectral intensity), and metabolite identification using mzCloud, mzVault, and MassList databases. Statistical standardization (QC-based Coefficient of Variation filtering, CV < 30%) and formula-based normalization were applied. Metabolites were annotated via KEGG, HMDB, and LIPIDMaps databases. Multivariate analyses (PCA, PLS-DA) and univariate t-tests (VIP > 1, $p < 0.05$, FC ≥ 2 or ≤ 0.5) identified differential metabolites, visualized by volcano plots and heatmaps (z-score normalized). Metabolic pathway enrichment ($x/n > y/N$, $p < 0.05$) and Pearson correlation analysis (corrplot package) were performed using R/Python. Instrument parameters: 3.5 kV spray voltage, 320°C capillary, 35 psi sheath gas, 10 L/min aux gas, 350°C heater (Want et al., 2013).

2.5 Combined transcriptome and metabolome analysis

Microbiome analysis enables the identification of structural and abundance variations in microbial communities, along with functional predictions or annotations, while metabolomics directly reflects the functional interactions between microbial communities and their hosts. These two approaches are complementary and indispensable. Integrating "microbiomics" with "metabolomics" allows a deeper understanding of how environmental microbial communities influence host or environmental metabolic states through microbial metabolism and host co-metabolism. In this study, we performed metabolomic profiling and metagenomic sequencing on samples, followed by correlation analysis between differential metabolites and microbial taxa (Hess et al., 2011).

Spearman rank correlation analysis measures the association between two variables using the Spearman correlation coefficient, with rank correlation tests to determine statistical significance. Unlike Pearson correlation, which assumes normal distribution, Spearman correlation relies on rank statistics and imposes no assumptions on data distribution. The Spearman coefficient ranges from -1 to 1, where positive/negative values indicate positive/negative correlations, and larger absolute values denote stronger associations ($|r| = 1$ indicates perfect correlation). Correlations were computed using the cor function in R, and significance (p-values) was assessed via the corPvalueStudent function from the WGCNA package (Ilhan et al., 2017).

3 Results

3.1 Diversity of rhizosphere bacterial and fungal communities among the different crop rotation and fertilization

ANOSIM and Adonis analyses of 16S (Supplementary Tables S2, S3) and ITS (Supplementary Tables S4, S5) revealed a convergent pattern of the four experimental groups. LEfSe analysis identified treatment-specific microbial biomarkers with linear discriminant analysis scores exceeding 4.0 (Figure 1). Bacterial communities in CK were characterized by Acidobacteria and Bacteroidia (LDA 4.2-4.5), whereas T3 exhibited significant enrichment of Thermoanaerobacter (LDA 4.8) associated with lignocellulose degradation. T1 elevated plant-associated Sphingomonas (LDA 4.3) to 10.4% compared to 6.3% in CK, while T2 enriched nitrogen-cycling Rhodanobacter (LDA 4.1) to 8.7% versus 5.2% in CK. Differential selection of Burkholderia was observed between T1 (LDA 3.9) and T2 (LDA 2.7), reflecting manure type-driven niche specialization. Fungal communities in CK were dominated by Absidia (LDA 4.0), T3

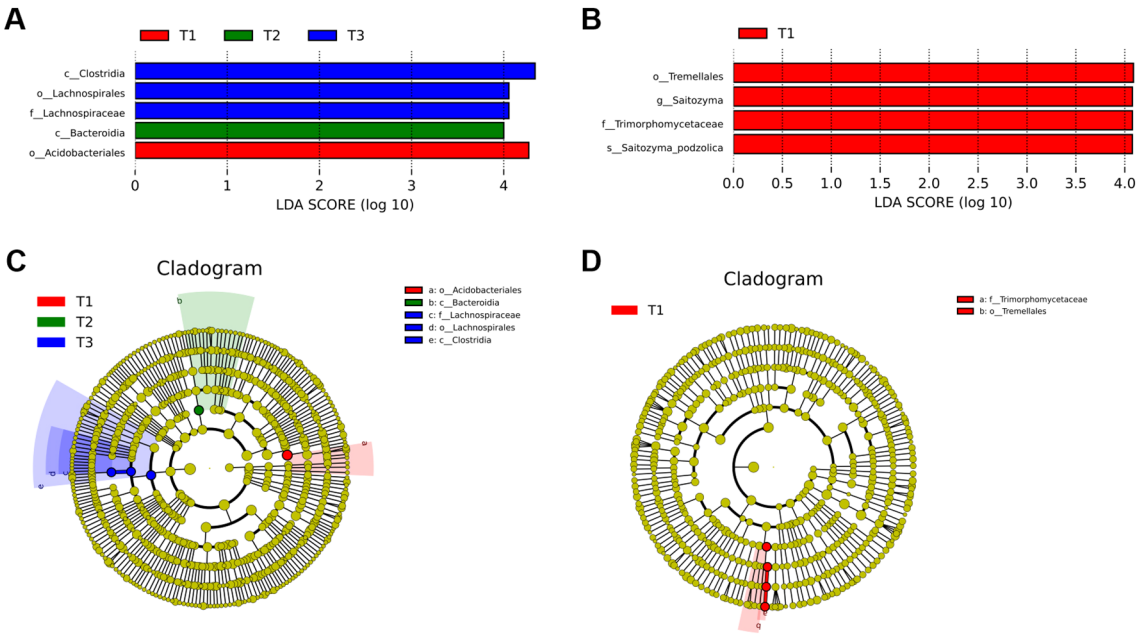


FIGURE 1
LEfSe analysis of microbial biomarkers across fertilization regimes. **(A)** LDA scores of bacterial biomarkers in T1 (red), T2 (green), and T3 (blue). **(B)** Fungal biomarkers with T1-specific enrichment. **(C)** Cladogram (phylum to family level) highlights T1-associated taxa (red), T2 (green), and T3 (blue); node size reflects relative abundance. **(D)** Phylogenetic clustering of T1-linked bacterial lineages (red gradient) and rare taxa (gray). Solid/dashed lines denote mean/median relative abundance; absent bars indicate undetectable biomarkers (LDA threshold >2, Kruskal-Wallis $p < 0.05$).

exhibited saprophytic Mortierella dominance (LDA 4.5), and T2 uniquely hosted mycorrhizal Entrophospora (LDA 4.2), which was undetected in other treatments. Fungal assemblages demonstrated lower β-diversity than bacterial counterparts, with Ascomycota consistently constituting 72-76% of communities across all groups. Critical functional divergences emerged wherein T3 preferentially recruited carbon-metabolizing taxa (Thermoanaerobacter, Mortierella), contrasting CK's Acidobacteria-Absidia profile, while T1 and T2 synergistically enriched nitrogen-cycling specialists (Sphingomonas, Rhodanobacter) and symbionts (Entrophospora in T2). Inter-treatment comparisons revealed T1's selection for Burkholderia versus T2's specialization toward Entrophospora, demonstrating crop-specific modulation of functional guilds. These results establish that fertilization strategies differentially reconfigure microbial consortia, with T3 enhancing organic matter mineralization capacity and T1/T2 promoting plant-microbe symbiotic networks.

Furthermore, Alpha diversity indices (Shannon and Chao1) revealed distinct microbial responses (Figure 2). For 16S, T3 had the highest bacterial diversity (Shannon: 9.5; Chao1: 2500) compared to CK (Shannon: 9.0; Chao1: 1500), driven by fertilizer cake-derived organic matter (Jiang et al., 2022). T1 and T2 showed moderate increases (Shannon: 9.3; Chao1: 2000), suggesting selective enrichment of copiotrophic taxa. For ITS, T3 also had the highest fungal diversity (Shannon: 5.5; Chao1: 600) compared to CK (Shannon: 3.5; Chao1: 300), reflecting fertilizer cake-induced proliferation of rare taxa like Mortierellomycota (Su et al., 2022). T1 and T2 had intermediate diversity (Shannon: 5.0–4.0; Chao1: 500–400), with T1 favoring Glomeromycota and T2 promoting Basidiomycota. T3 significantly increased microbial diversity over CK (bacterial Chao1: +66.7%; fungal Chao1: +100%), while T1 and T2 showed moderate increases (bacterial Chao1: +33.3%; fungal Chao1: +66.7% in T1, +33.3% in T2). T1 prioritized symbiotic functions, while T2 favored decomposition. Between T1 and T2, bacterial diversity was similar (Chao1: 2000), but T2 had lower fungal diversity (Chao1: 400 vs. 500 in T1), likely due to rape's recalcitrant lignin enriching specialized degraders.

The phylum-level composition of bacterial (16S) and fungal (ITS) communities showed distinct responses to treatments (Figure 3). In 16S analysis, CK was dominated by Proteobacteria and Actinobacteria. T3 had increased Firmicutes and Bacteroidota, linked to enhanced organic matter decomposition from high-carbon fertilizer cake (Rampadarath et al., 2018). T1 showed reduced Proteobacteria but elevated Acidobacteria, likely due to *V. myuros*-induced soil acidification (Kalam et al., 2020). T2 uniquely enriched Chloroflexi, potentially tied to recalcitrant lignin from rape residues (Atiwesh et al., 2022). For ITS analysis, Ascomycota dominated all groups. T3 had reduced Basidiomycota, possibly due to fertilizer cake suppressing ligninolytic fungi (Atiwesh et al., 2022). T1 showed a striking increase in Glomeromycota, consistent with *V. myuros*-enhanced arbuscular mycorrhizal symbiosis (Sawers et al., 2018). T2 had higher Basidiomycota, likely supporting rape lignocellulose degradation (Atiwesh et al., 2022). Bacterial communities in T1 favored oligotrophic Acidobacteria, while T2 enriched copiotrophic Chloroflexi. Fungal communities in T1 were dominated by Glomeromycota, contrasting with Basidiomycota specialization in T2,

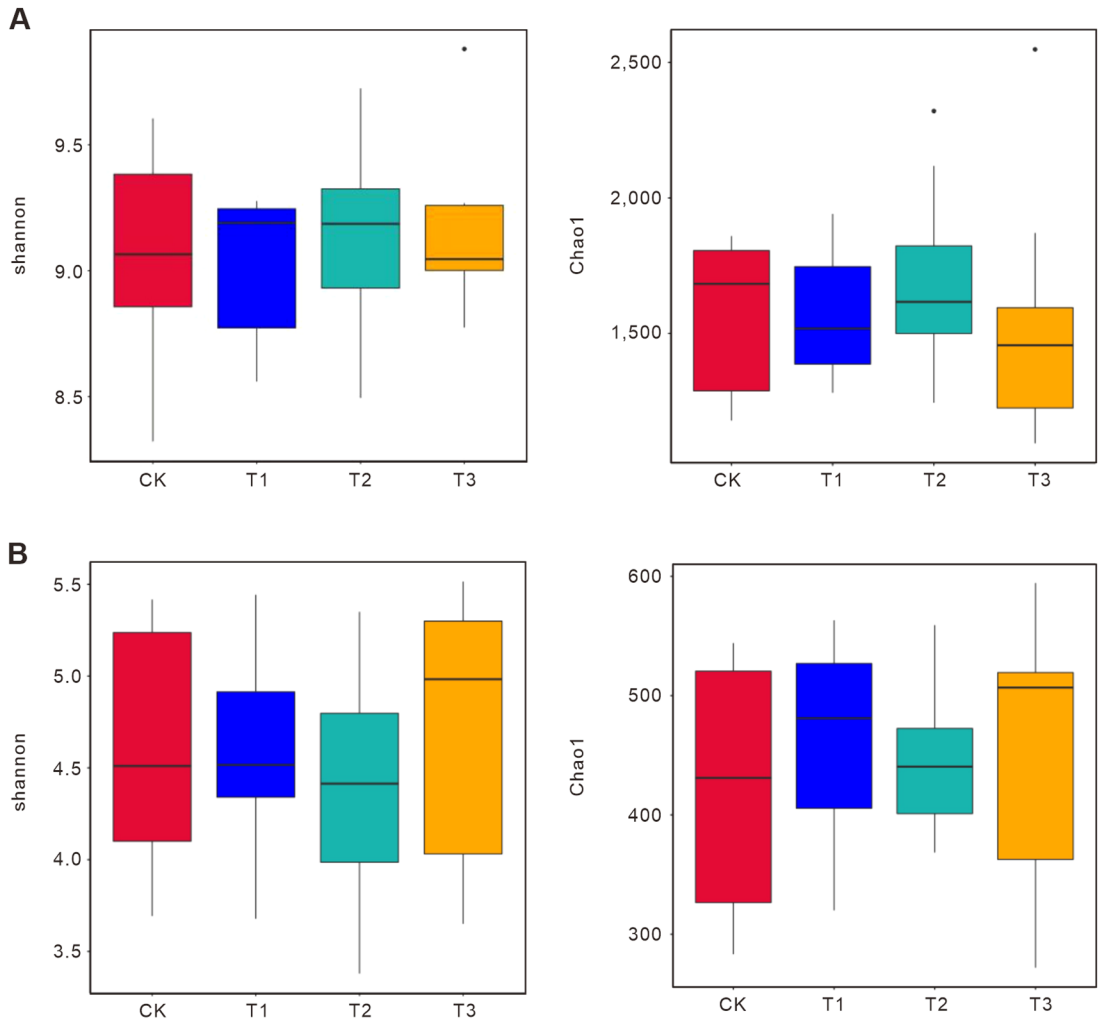


FIGURE 2
Alpha diversity indices (Shannon and Chao1) of bacterial and fungal communities across treatments. Alpha diversity indices of bacterial **(A)** and fungal **(B)** reflect species richness and evenness within samples. Higher index values indicate greater microbial community complexity. Statistical significance tests between groups can quickly identify groups with significantly increased or decreased diversity for further biological analysis. Boxplots visually compare diversity across groups. The Shannon index accounts for total taxa and their proportional abundances, with higher values indicating greater diversity and even species distribution. The Chao1 index estimates total species richness in a community, incorporating data from species with abundance 1 and 2 to better reflect low-abundance taxa. y-axis: index values; x-axis: treatment groups. "*" indicate statistically significant differences between corresponding groups. CK: Conventional fertilization. T1: Rotation + Intercropping group. T2: Rotation group. T3: Cake fertilizer group.

reflecting differential carbon utilization between *V. myuros* and rape cultivation (Zhou et al., 2018). These results highlight treatment-specific impacts: T3 enhanced decomposition, T1 promoted symbiosis, and T2 favored recalcitrant carbon degradation.

Barplots of genus level demonstrated distinct treatment-specific microbial responses. T3 significantly enriched Fusarium, with bacterial abundance reaching 14.8% compared to 3.2% in control (CK) and fungal abundance at 18.3% versus 6.5% in CK, alongside Humicola at 12.1% relative to 3.9% in CK, indicating a selective enhancement of carbon-degrading taxa. T1 elevated Sphingomonas to 10.4% from 6.3% in CK and increased Entolomataceae to 9.4%, while T2 upregulated Rhodanobacter to 8.7% compared to 5.2% in CK and Sporophila to 10.2%, reflecting functional niche differentiation driven by green manure specificity. Bacterial communities exhibited stronger treatment-associated divergence, as exemplified by Burkholderia abundance (7.1% in T1 versus 4.9% in T2), whereas fungal communities maintained structural consistency with Ascomycota persistently dominating across all treatments at 72–76% relative abundance.

Species-level barplots demonstrated distinct treatment effects. Bacterial communities in CK predominantly harbored Acidobacteria and Roseburia, whereas T3 exhibited significant enrichment of thermophilic Thermoanaerobacter taxa linked to carbon metabolism. T1 elevated plant growth-promoting Sphingomonas to 10.4% compared to 6.3% in CK, while T2 enhanced nitrogen-cycling Rhodanobacter to 8.7% versus 5.2% in CK, concurrent with differential Burkholderia abundance between T1 (7.1%) and T2 (4.9%). Fungal profiles revealed CK dominated by Absidia, T3 characterized by saprophytic Mortierella, and T2 uniquely hosting mycorrhizal Entrophospora at 9.4% abundance,

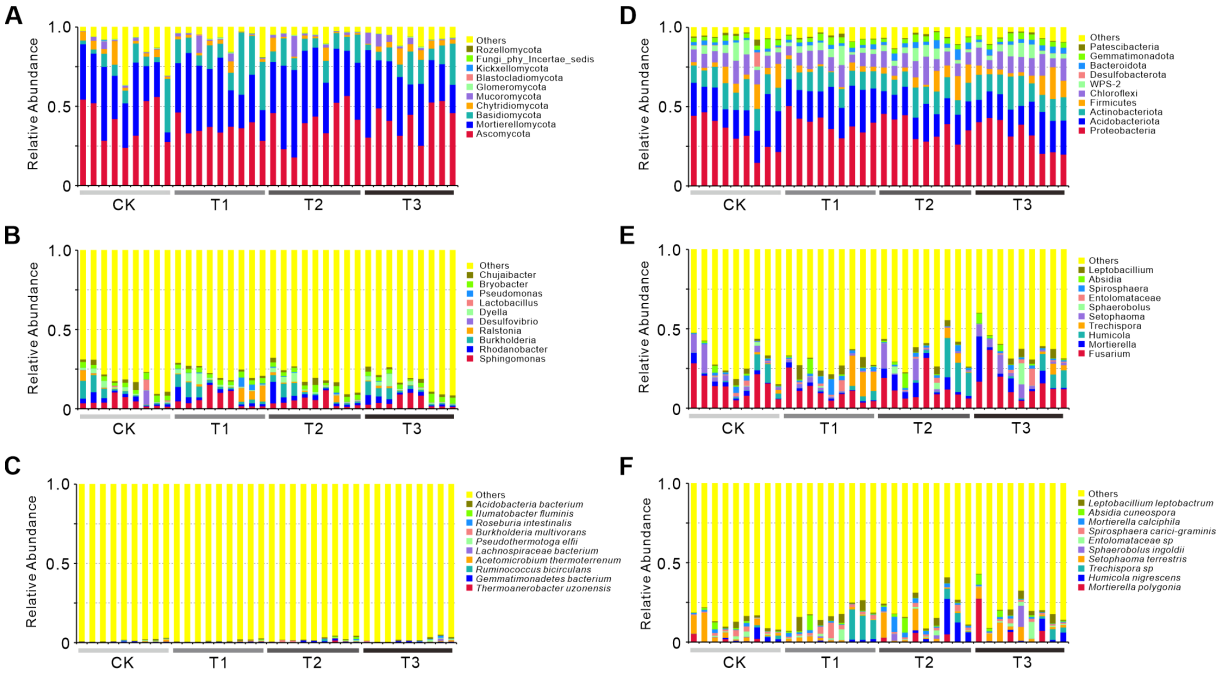


FIGURE 3
Relative abundance of bacterial and fungal communities at the phylum, genus, and species levels across treatments. Relative abundance of bacterial **(A, B, C)** and fungal **(D, E, F)** communities at the phylum **(A, D)**, genus **(B, E)**, and species **(C, F)** levels across treatments (CK, T1, T2, T3). Bar charts display the top 10 taxa by relative abundance at the phylum, genus, and species levels, with remaining taxa grouped as 'Others'. Colors distinguish microbial taxa, while the x-axis indicates sample names and the y-axis shows relative abundance (%). Each bar aggregates low-abundance species into the 'Others' category, calculated by summing relative abundances beyond the top 10 ranked taxa. CK: Conventional fertilization. T1: Rotation + Intercropping group. T2: Rotation group. T3: Cake fertilizer group.

with fungal assemblages displaying reduced inter-treatment variability relative to bacterial counterparts as Ascomycota consistently constituted 72–76% of total communities. Critical functional divergences emerged wherein T3 prioritized carbon-metabolizing taxa, whereas T1 and T2 enriched nitrogen cyclers and symbionts, reflecting crop-specific microbial selection dynamics.

3.2 PCA and KEGG pathway analysis of tobacco rhizosphere soil metabolites

Principal component analysis (PCA) of four experimental and QC samples revealed distinct scores for PC1 and PC2 (Supplementary Figure S1). The concentration of QC samples indicates that the test process was highly stable (Bokulich et al., 2013). Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis of the annotated metabolites identified six major functional categories (first-level classification): Cellular processes, drug development, environmental information processing, genetic information processing, metabolism, and organismal systems. Among the secondary classification units, the top three categories were global and overview maps, amino acid metabolism, and lipid metabolism (Figure 4). These pathways, particularly those related to global and overview maps, amino acid metabolism, and lipid metabolism, are crucial for understanding the tobacco soil metabolome (Shi et al., 2024). The global and overview maps pathway provides a comprehensive view of tobacco soil metabolism through the KEGG atlas, which includes global metabolic maps, data visualization, and auxiliary pathway analysis (Qiao et al., 2023). This helps in understanding the interrelationships, changes, and mechanisms of tobacco soil metabolites. Microorganisms influence amino acid content and composition in the soil by affecting amino acid uptake and utilization, which is further influenced by microbial community structure, soil environmental factors (such as pH and carbon content), and the feedback from metabolic products (Babalola et al., 2021).

3.3 Microbial co-occurrence networks reveal treatment-specific interaction patterns

Spearman-based co-occurrence networks revealed treatment-specific modulation of bacterial (16S) and fungal (ITS) interactions (Figure 5). For 16S data, T3 exhibited the highest connectivity (edges: 68 vs. CK: 41), dominated by Firmicutes and Bacteroidota

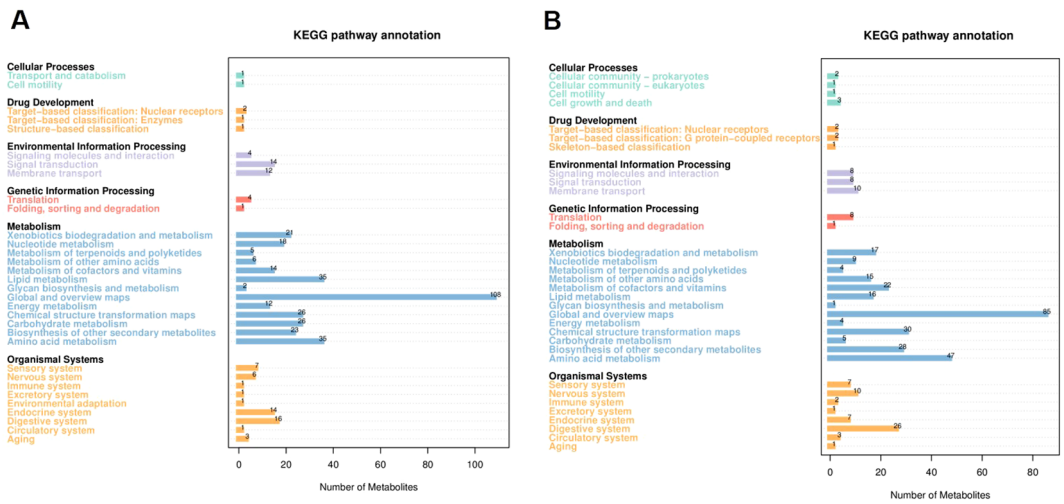


FIGURE 4
The annotation of negative and positive metabolites within the KEGG pathways. The horizontal coordinates negative **(A)** and positive **(B)** represent the number of metabolites, and the vertical coordinates represent the annotated KEGG pathways. This figure provides a comprehensive overview of the distribution of metabolites across various KEGG pathways at the secondary classification level. Each bar in the bar chart corresponds to a specific secondary-classified KEGG pathway, with the height of the bar indicating the number of metabolites associated with that particular pathway.

with 78% positive correlations, indicating fertilizer cake-enhanced synergistic decomposition. CK showed sparse connections among Proteobacteria, while T1 formed modular clusters between Acidobacteria and Proteobacteria (edges: 55), reflecting niche partitioning driven by *V. myuros*. T2 displayed fragmented networks centered on Chloroflexi, with 49% negative correlations, suggesting lignin-driven competitive exclusion.

For ITS data, T3 featured reduced Basidiomycota interactions but strengthened positive correlations between Mortierellomycota and Ascomycota (edges: 45 vs. CK: 28), aligning with fertilizer cake-mediated lipid metabolism (Fallah et al., 2023). T1 showed Glomeromycota hubs (degree: 8) with 89% positive correlations, supporting *V. myuros*-induced mycorrhizal symbiosis, while T2 centered on Basidiomycota-Chytridiomycota co-dependencies (edges: 32), driven by rape lignocellulose degradation.

Comparisons between treatments further highlighted distinct microbial dynamics. T3 bacterial edge density increased by 65.9%, and fungal positive correlations rose by 24.1% compared to CK, underscoring fertilizer cake's role in fostering cooperative guilds. T1 exhibited higher bacterial modularity (0.46 vs. CK: 0.32) and expanded fungal Glomeromycota nodes (6 vs. CK: 2), reflecting intercropping-mediated mutualism (Wang et al., 2016). In contrast, T2 showed dominant bacterial negative correlations (48.7% vs. CK: 31.5%) and fungal Basidiomycota specialization, indicating rape residue-driven competition. Between T1 and T2, T1 prioritized Acidobacteria symbiosis (edges: 14 vs. T2: 5), while T2 emphasized Basidiomycota-Chytridiomycota decomposer co-occurrence (edges: 9 vs. T1: 2). These results demonstrate that carbon source quality (fertilizer cake vs. plant residues) and planting regimes (intercropping vs. rotation) reconfigure microbial interaction strategies, fostering distinct functional guilds and ecological dynamics.

3.4 Lipid metabolite profiling in tobacco rhizosphere soil

Lipid metabolites were annotated using the lipid maps database, revealing a total of 86 down-regulated lipid metabolites (Figure 6). The top three were 23 types of Fatty Acids and Conjugates [FA01], 17 types of Eicosanoids [FA03], and 9 types of Glycerophosphoethanolamines [GP02]. Additionally, 45 up-regulated lipid metabolites were identified, with the top three being 7 types of fatty acids and conjugates [FA01], 7 types of flavonoids [PK12], and 6 types of steroids [ST02].

Fatty acids and conjugates [FA01] can indirectly influence tobacco growth by affecting the structure of soil microbial communities and participating in nutrient cycling. Eicosanoids [FA03] serve as plant-microbe signaling molecules in tobacco soil. They facilitate the aggregation of beneficial microorganisms and play a critical role in tobacco defense mechanisms by activating defense gene expression. In conclusion, lipid metabolite profiling in tobacco soil revealed significant changes, with down-regulated fatty acids and eicosanoids influencing microbial community structure, nutrient cycling, and plant defense, while up-regulated metabolites, such as flavonoids and steroids, further contributed to soil ecosystem stability, providing a clear overview of the lipidomic changes that underpin soil health and tobacco growth.

3.5 Pathway enrichment analysis of differential metabolites

To investigate the potential effects of differential metabolites, pathway enrichment analysis was performed using the KEGG database (Figure 7). Compared to the control (CK), T1 treatment

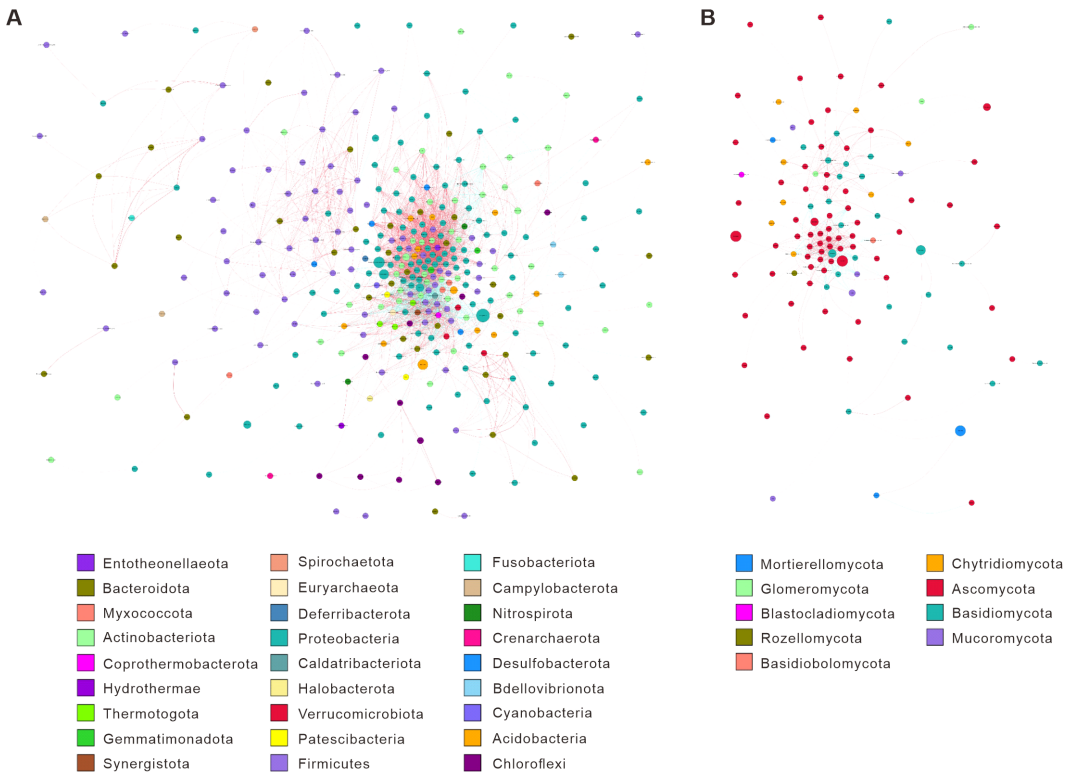


FIGURE 5
Microbial interaction networks based on Spearman correlations for 16S and ITS sequencing across treatments. The network diagram was generated by calculating Spearman correlation coefficients for 16S **(A)** and ITS **(B)** sequencing across all samples to form a species correlation matrix. Filtering conditions were applied: removing correlations with coefficients <0.6; excluding self-connections; discarding links where node abundance was <0.005%. Nodes represent genera (size indicates average relative abundance; colors denote phylum-level taxonomy). Edges reflect species interactions: thickness corresponds to the absolute correlation coefficient, while colors indicate positive (red) or negative (blue) correlations.

resulted in significantly down-regulated metabolites enriched in pathways such as the biosynthesis of phenylpropanoids, biosynthesis of amino acids, and the biosynthesis of alkaloids derived from the shikimate pathway. In T2 treatment, down-regulated metabolites were enriched in the biosynthesis of phenylpropanoids, biosynthesis of amino acids, and carbon metabolism. The changes in phenylpropanoid and amino acid biosynthesis were similar to those observed in T1. For T3 treatment, down-regulated metabolites were significantly enriched in pathways such as tryptophan metabolism and microbial metabolism in diverse environments. Up-regulated metabolites in T3 were enriched in pathways related to microbial metabolism, biosynthesis of secondary metabolites, and plant secondary metabolite biosynthesis.

To explore the differences of soil metabolites in different rotation groups, we also compared T1 and T2. Compared to T1, T2 treatment showed significant down-regulation of metabolites enriched in pathways such as Tryptophan Metabolism, the Phosphotransferase System (PTS), and Bile Secretion. Up-regulated metabolites in T2 were enriched in pathways such as Tropane, Piperidine, and Pyridine Alkaloid Biosynthesis, Microbial Metabolism in Diverse Environments, and Alkaloid Biosynthesis from Ornithine, Lysine, and Nicotinic Acid.

The total differential metabolite heatmap visually reveals the impact of different fertilization treatments on soil microbial-metabolic interaction networks (Supplementary Figure S2). In T1 treatment, 37 metabolites were significantly down-regulated compared to CK, including 2'-Deoxycytidine 5'-diphosphate (dCDP), 3-Hydroxybenzoic acid, Pyrithioxin, and Sodium cholate. Meanwhile, 67 metabolites were up-regulated, including 1,3-Dihydro-1,3,3-trimethyl-2H-indol-2-ylidene acetaldehyde, N-Acetyl-DL-phenylalanine, and 2-Isobutyl-3-methoxypyrazine.

In T2 treatment, 35 metabolites showed significant down-regulation compared to CK, including 1-(4-Methoxyphenyl)-2-propanone, 16-Hydroxyhexadecanoic acid, and 4-Nitrophenol. In contrast, 31 metabolites were up-regulated, including 5-phenyl-2,3-dihydro-1H-1,4-benzodiazepin-2-one, Isohomovanillic acid, and Kaempferol.

In T3 treatment, 24 metabolites were significantly down-regulated compared to CK, including Hexadecanedioic acid, Delta-Tridecalactone, and Glycocholic acid. Additionally, 39 metabolites were up-regulated, including 4-(3-methoxy-5,6-dihydrobenzo[c]acridin-7-yl) morpholine, 1,4-dihydroxyheptadec-16-en-2-ylacetate, and Pyridoxamine.

When comparing T2 with T1, 48 metabolites were significantly down-regulated in T2, including 1-(4-Methoxyphenyl)-2-

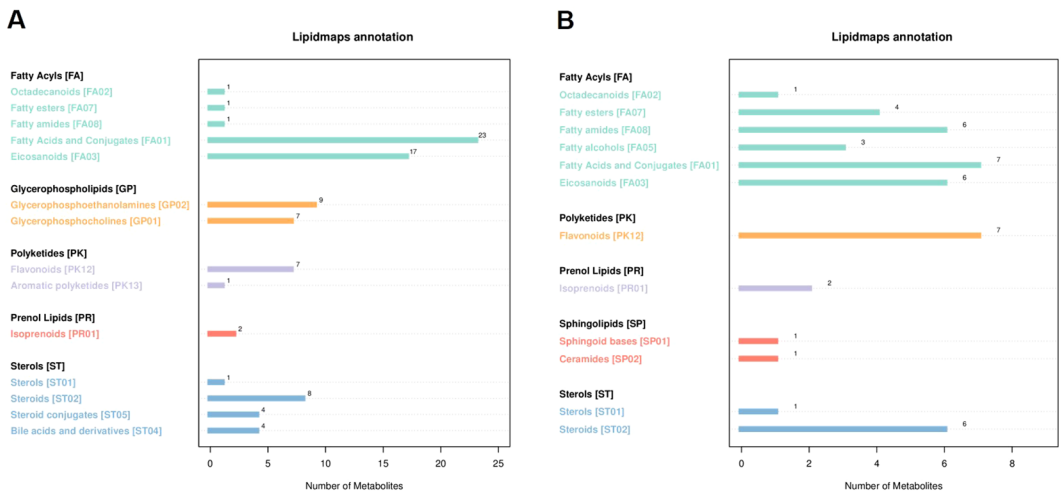


FIGURE 6
The number of metabolites, and the ordinate represents the lipid classification of LIPID MAPS annotated. **(A)** negative lipid metabolites. **(B)** positive lipid metabolites. This figure shows the number of (lipid) metabolites corresponding to the main level classification (Main Class) under the eight major lipid classifications (Category) in LIPID MAPS. The x-axis indicates the number of metabolites, while the y-axis lists the lipid classes. The bar lengths in each panel reflect the metabolite counts for each lipid class.

propanone, 16-Hydroxyhexadecanoic acid, and N-Formylkynurenine. Conversely, 72 metabolites were up-regulated in T2, including 1,3-Dihydro-1,3,3-trimethyl-2H-indol-2-ylidene acetaldehyde, N-Acetyl-DL-phenylalanine, and (+/-)5(6)-EET Ethanolamide. Overall, the analysis revealed that different tobacco planting treatments significantly affected key metabolic pathways, including phenylpropanoid and amino acid biosynthesis, carbon metabolism, and microbial interactions, highlighting the functional shifts caused by different agricultural practices.

3.6 Gene abundance in soil microorganisms across treatments using 16S sequencing analysis

The PCA plots of the four different experimental groups for 16S sequencing show that the experimental results of each group are relatively clustered, indicating that the experimental results are relatively stable (Supplementary Figure S3). The cluster heatmap from 16S sequencing of soil microorganisms revealed distinct differences in gene abundance across treatments (Figure 8). Compared to the CK, T1 treatment showed higher abundance of genes related to human diseases and environmental information processing (Level 1), such as transcription, signal transduction, and infectious and neurodegenerative diseases (Level 2). In T2, genes associated with genetic information processing and cancers (Level 2) showed higher abundance. Genetic information processing includes key cellular processes such as DNA replication, transcription, and translation. The elevated abundance of these genes suggests that soil microorganisms in T2 are actively replicating and expressing their genetic materials to adapt to the environmental conditions. In T3, genes related to metabolism and organismal systems (Level 1), including Metabolism of terpenoids and polyketides and amino acid metabolism (Level 2), were more abundant. Metabolism is a fundamental process for maintaining life, and the high abundance of metabolism-related genes in T3 indicates that the microorganisms are actively engaged in substance metabolism, obtaining energy and nutrients from the environment (McClure et al., 2020).

To sum up, 16S sequencing analysis revealed that different tobacco planting treatments influenced microbial gene abundance, with T1 showing increased transcription and signal transduction activity, T2 promoting genetic information processing, and T3 enhancing metabolic pathways related to nutrient acquisition.

3.7 ITS sequencing results of soil microorganisms across treatments

The PCA plots of the four distinct experimental groups for ITS sequencing demonstrate that the experimental outcomes of each group are relatively clustered, suggesting relatively stable experimental results (Supplementary Figure S4). The ITS sequencing results revealed significant differences in gene abundance across treatments (Figure 8). Compared to the CK, T1 treatment showed higher abundance of genes related to arbuscular mycorrhizal fungi, fungal parasites, undelined saprotrophs, wood saprotrophs, dung saprotrophs, endophytes, and plant pathogens. In T2, genes associated with undefined saprotrophs, wood saprotrophs, endophytes, and animal pathogens showed higher abundance. In T3, genes related to dung saprotrophs, undefined saprotrophs, and animal endosymbionts were more abundant.

In brief, the ITS sequencing analysis revealed distinct shifts in fungal gene abundance across treatments, with T1 showing increased levels of arbuscular mycorrhizal fungi and saprotrophic fungi, T2 enriching for wood saprotrophs and animal pathogens,

10.3389/fpls.2025.1595870

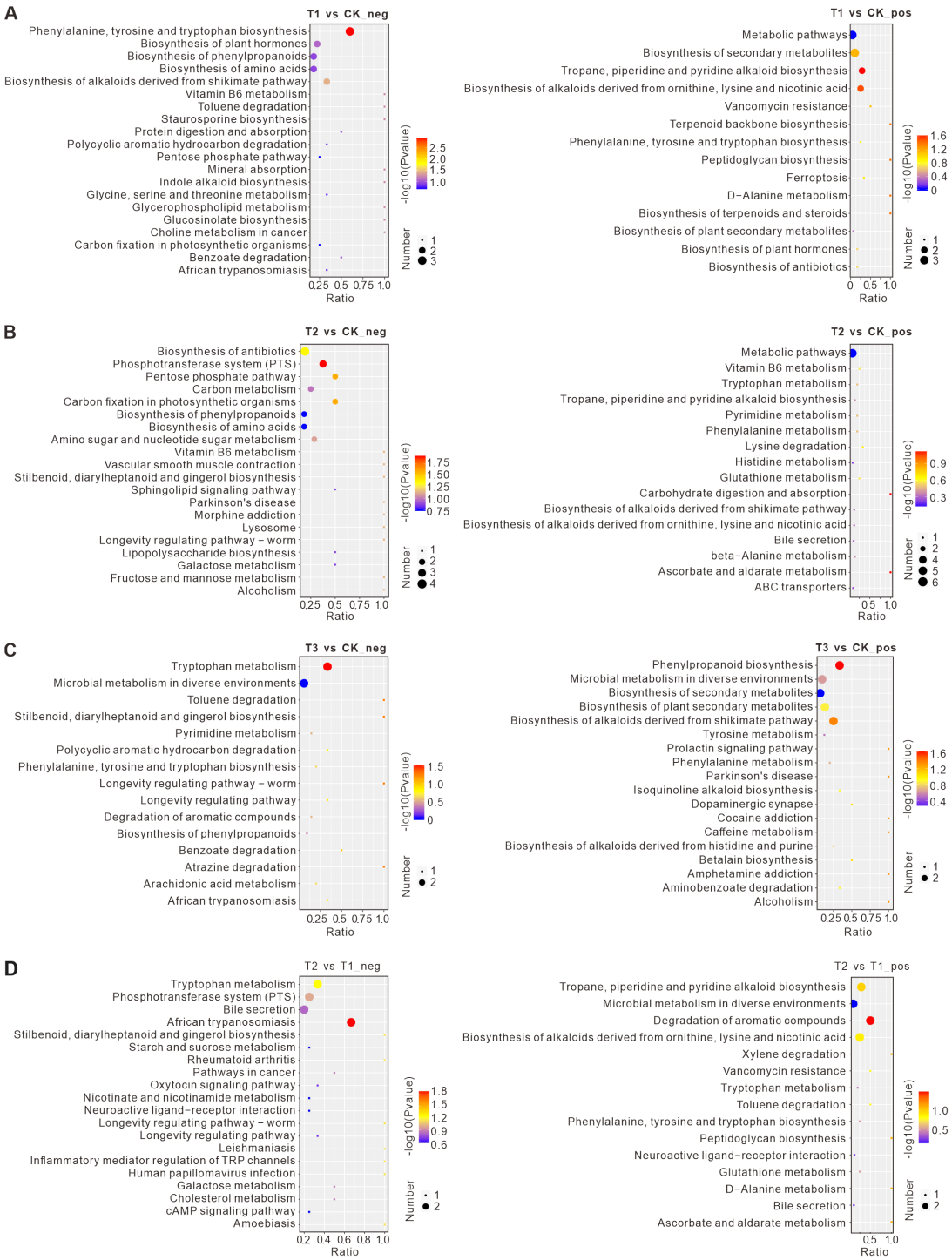


FIGURE 7
The top 20 pathways according to the KEGG enrichment results. **(A)** Negative and positive metabolites in the comparison between T1 and CK. **(B)** Negative and positive metabolites in the comparison between T2 and CK. **(C)** Negative and positive metabolites in the comparison between T1 and T2. **(D)** Negative and positive metabolites in the comparison between T3 and CK. The abscissa is x/y (the number of differential metabolites in the corresponding metabolic pathway/the total number of metabolites identified in the pathway). The larger the value, the higher the enrichment degree of differential metabolites in the pathway. The ordinate is the name of the KEGG pathway. The color of the point represents the P-value of the hypergeometric test. The smaller the value, the greater the reliability of the test and the more statistically significant. The size of the point represents the number of differential metabolites in the corresponding pathway. The larger the size, the more differential metabolites there are in the pathway. CK: Conventional fertilization, T1: Rotation + Intercropping group, T2: Rotation group, T3: Cake fertilizer group.

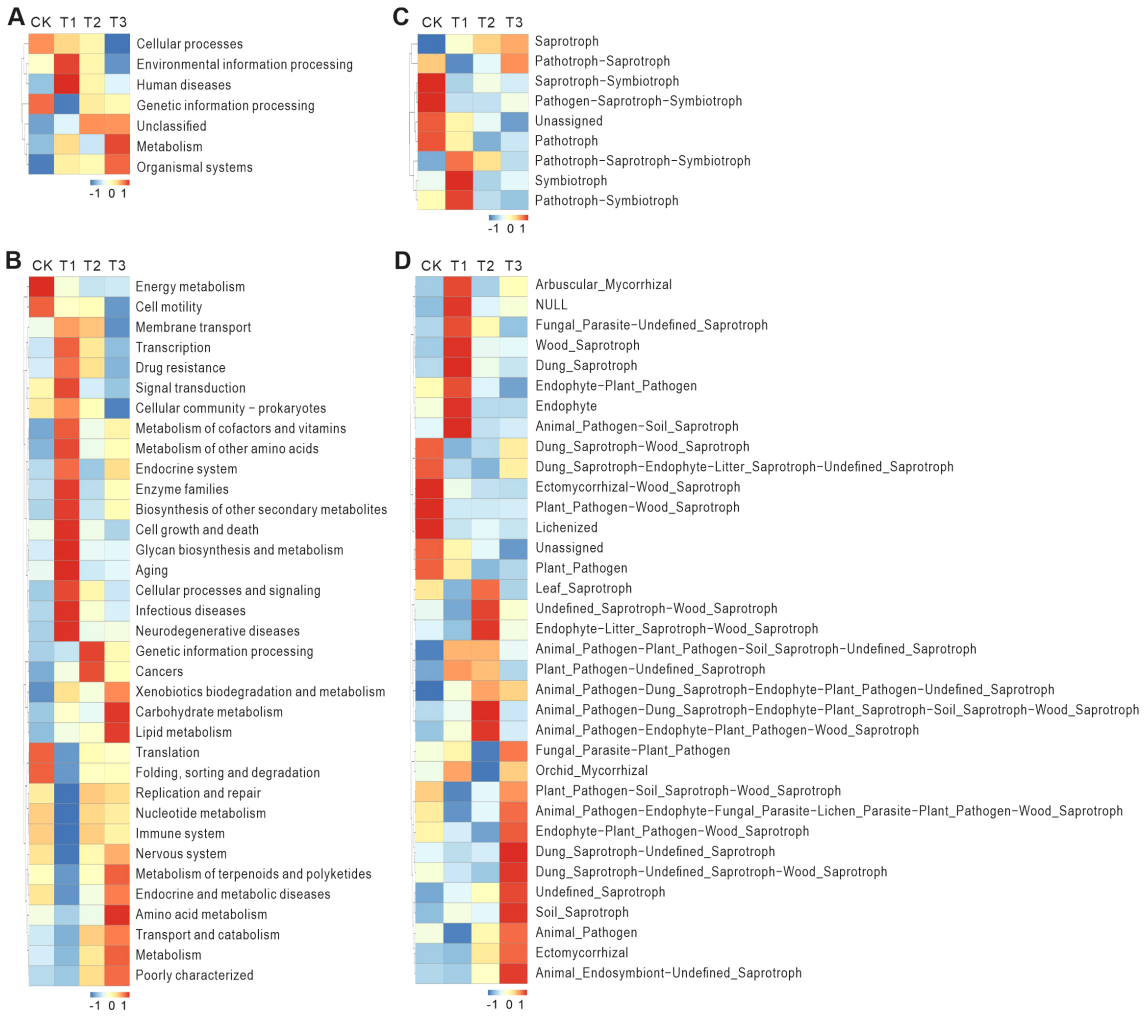


FIGURE 8
16S and ITS sequencing predicts functional annotation clustering heatmaps, displaying the same species heatmap. This heatmap is completed using Tax4Fun and FunGuild. Tax4Fun is an R package based on the 16S Silva database for functional prediction of environmental samples such as gut and soil. Based on the functional annotations and abundance information of samples in the database, the top 35 functions in terms of abundance and their abundance information in each sample are selected to draw the heatmap, and clustering is performed from different functional levels. FunGuild is an amplicon analysis based on ITS that can obtain the classification and abundance information of fungal species present in the environment, and can correspond to the ecological functions of fungi based on their species classification. Based on the functional annotations and abundance information of samples in the database, the top 35 functions in terms of abundance and their abundance information in each experimental group are selected to draw the heatmap, and clustering is performed from the perspective of functional differences. **(A)** The functional prediction analysis results of Level1 database. **(B)** The functional prediction analysis results of Level2 database. **(C)** The functional prediction analysis results of the mode database. **(D)** The functional prediction analysis results of the guild database. Horizontally represents experimental groups (CK: Conventional fertilization, T1: Rotation + Intercropping group, T2: Rotation group, T3: Cake fertilizer group), and vertically represents functional annotation information, with the clustering tree on the left side of the heatmap being the functional clustering tree; the corresponding values in the heatmap are the Z-scores obtained after standardizing the relative abundance of each row of functions, that is, the Z-score of a sample in a certain category is the difference between the relative abundance of the sample in that category and the average relative abundance of all samples in that category, divided by the standard deviation of all samples in that category.

and T3 promoting dung saprotrophs and animal symbionts, reflecting complex interactions between fungi, plants, and soil conditions.

3.8 Link between microbes and metabolites

For combined 16S sequencing and metabolome analysis,the heatmap delineates significant Spearman correlations between microbial taxa and the top 20 variable importance in projection (VIP)-ranked metabolites across fertilization regimes (Figure 9). In comparative analyses between CK and T1, Actinospica belonging to Actinobacteria demonstrated a robust positive correlation with FR01.1801-Hydrogen, a carbon-nitrogen cycling-associated compound, exhibiting a correlation coefficient of 0.75 at a significance level of $p < 0.05$. This aligns with the role of T1 in enhancing organic matter turnover. Conversely, Rudena from Proteobacteria displayed a negative correlation with Copoxyoriso (a), a phenolic derivative, showing a coefficient of -0.50 ($p < 0.05$),

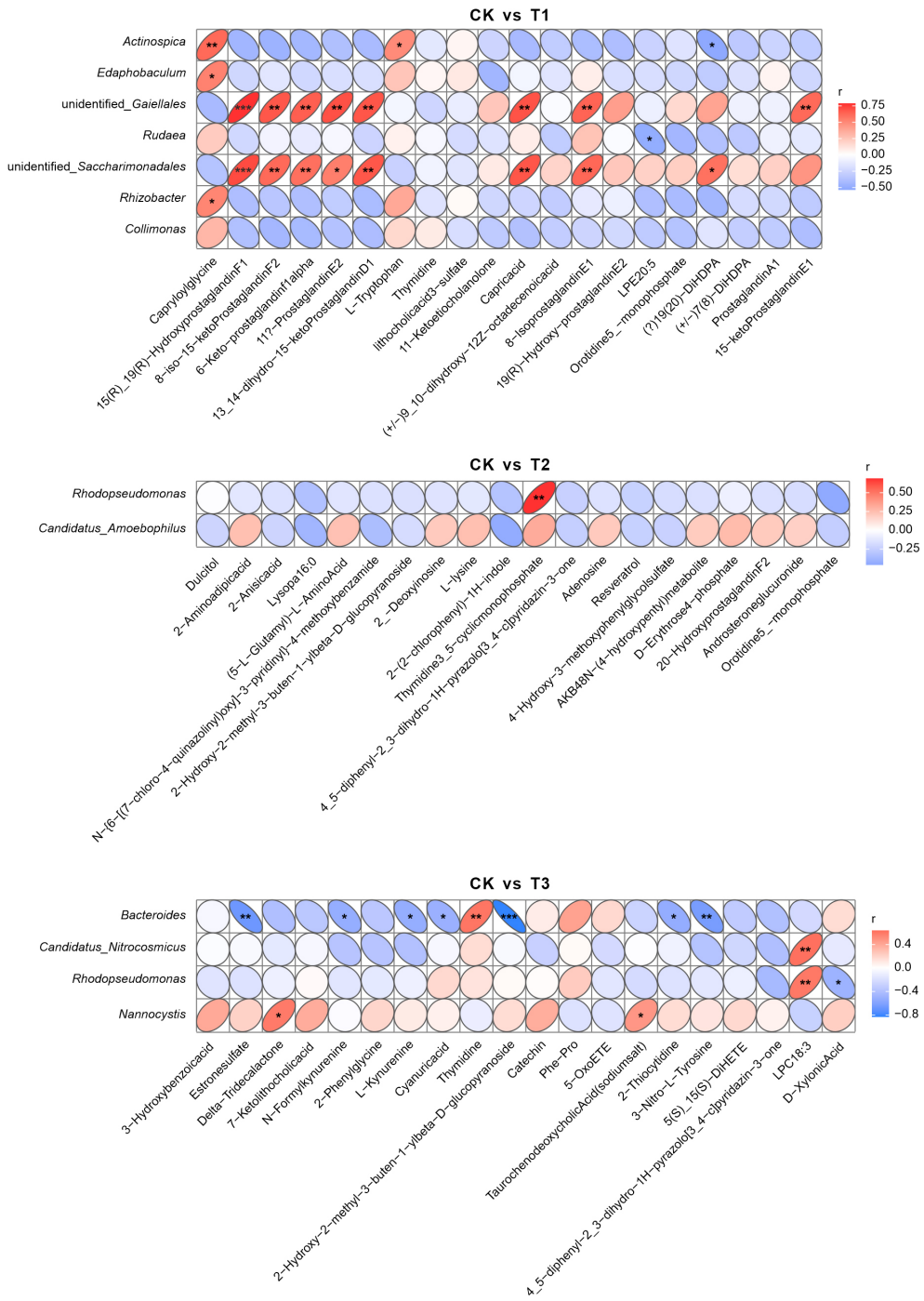


FIGURE 9
Heatmaps of Spearman correlations between microbial taxa and VIP-selected differential metabolites. The heatmaps demonstrate microbial genera/species (y-axis) versus metabolites (x-axis) across pairwise comparisons: CK vs T1, CK vs T2, and CK vs T3. Red/blue ellipses indicate positive/negative correlations (|r| ≥ 0.25, p < 0.05), with ellipse thinness inversely proportional to |r| magnitude. Blank cells denote non-significant associations (p ≥ 0.05). CK: Conventional fertilization. T1: Rotation + Intercropping group. T2: Rotation group. T3: Cake fertilizer group. *P<0.05; **P<0.01; ***P<0.001.

indicative of suppressed oxidative stress-related metabolite synthesis under T1.Comparisons between CK and T2 revealed moderate positive correlations (r = 0.50, p < 0.05) between Phodopsseudomonas (Proteobacteria) and 2-Antropid(Ri), a nitrogen-containing alkaloid, consistent with T2 promoting nitrogen-metabolizing taxa through rape rotation. In CK versus T3 analyses, Candidatus_Attracolbidon exhibited a weak negative association (r = -0.40, p < 0.05) with D-Xivlenol, a lignin-derived compound, suggesting fermented cake amendments may attenuate lignocellulose degradation activity.These results demonstrate that T1 and T2 strengthen microbial-metabolite interactions associated with nutrient cycling, whereas T3 fosters taxon-specific but

functionally weaker associations, reflecting divergent mechanisms by which organic amendments reconfigure rhizosphere metabolic networks.

For combined ITS sequencing and metabolome analysis, the heatmap delineates significant Spearman correlations between fungal taxa and the top 20 variable importance in projection (VIP)-ranked metabolites across fertilization regimes (Supplementary Figure S5). Comparative analysis between CK and T1 revealed that the fungal genus Saitozyma exhibited a robust positive correlation ($r = 0.50$, $p < 0.05$) with prostaglandin F1, a plant hormone-like metabolite implicated in stress signaling pathways, consistent with T1 promoting stress-responsive fungal symbionts. Conversely, Nectria demonstrated a negative correlation ($r = -0.25$, $p < 0.05$) with 8-iso-15-ketoProstaglandin F1, an oxidative stress-associated prostaglandin derivative, indicative of T1-mediated suppression of fungal oxidative stress pathways.In CK versus T2 comparisons, the taxon Oplidiomycota_gen_incertae_sedis displayed moderate positive correlations with resveratrol ($r = 0.40$, $p < 0.05$), a phytoalexin exhibiting antioxidant properties, alongside negative correlations with androsterone glucuronide ($r = -0.40$, $p < 0.05$), a steroid metabolite. This dual correlation pattern suggests T2 preferentially enriches fungal taxa involved in secondary metabolite biosynthesis while reducing steroid accumulation.These results demonstrate that T1 strengthens fungal interactions with stress-related metabolic networks, whereas T2 drives fungal community specialization toward antioxidant synthesis, reflecting distinct functional adaptation mechanisms to organic amendment strategies.

4 Discussion

This study presents comprehensive insights into the effects of different tobacco planting systems on soil metabolites, microbial communities, and their functional roles in the rhizosphere. The results suggest that crop rotation and intercropping with species such as *V. myuros* and rape can significantly alter soil metabolite profiles and microbial diversity, with important implications for soil health, plant growth, and pest management (Liu et al., 2024b).

PCA and KEGG pathway analyses revealed notable shifts in tobacco soil metabolomes under different treatments (Magoč and Salzberg, 2011). The pathways related to amino acid metabolism, lipid metabolism, and global metabolic maps were particularly prominent, highlighting the complex biochemical interconnections that govern nutrient cycling and plant health (Sun et al., 2023b). Fatty acids and conjugates [FA01] down-regulated in T1 were particularly linked to changes in microbial cell membrane integrity, which may have influenced microbial diversity and nutrient cycling processes. Changes in these metabolites can alter microbial cell membrane characteristics, thus impacting microbial populations (Wu et al., 2024a). Their metabolic products contribute to the cycling of key elements, such as carbon and nitrogen, ultimately affecting the nutrient availability and growth of tobacco plants (Zhang et al., 2024). Up-regulation of flavonoids and steroids in T1 may have contributed to improved soil stability and plant defense mechanisms, underscoring the intricate feedback between tobacco plants and their microbial environment (Zhuang et al., 2024). Eicosanoids [FA03] serve as plant-microbe signaling molecules in tobacco soil. They facilitate the aggregation of beneficial microorganisms and play a critical role in tobacco defense mechanisms by activating defense gene expression. These metabolites also contribute to soil ecosystem stability by reducing the accumulation of pathogenic bacteria and maintaining microbial diversity (Fallah et al., 2023). Changes in tryptophan metabolism are linked to neurotransmitter synthesis and physiological regulation, while the PTS plays a key role in substance transport and energy metabolism (Wang et al., 2023). Alterations in bile secretion pathways may affect substance metabolism and detoxification (Liu et al., 2024c).

In addition, lipid metabolism pathways, which interact with other pathways such as fatty acid metabolism, PPAR signaling, and glyceride metabolism, also play a significant role in tobacco soil metabolomics (Lin et al., 2022). These pathways impact the soil metabolites by altering microbial community structure, regulating gene expression, and participating in signal transduction (Kumawat et al., 2022). Moreover, continuous tobacco cropping and the use of different tobacco varieties affected rhizosphere microorganisms and metabolites, indirectly linking lipid metabolism pathways to other biological processes (Chen et al., 2024b). Altogether, these findings emphasize the complex relationship between tobacco cultivation, soil metabolites, and microbial diversity.Phenylpropanoids play critical roles in plant growth, stress resistance, and development (Wang et al., 2020). Amino acids, as fundamental building blocks of proteins, are crucial for protein synthesis and metabolism, influencing overall physiological functions (Lei et al., 2017). The alkaloid biosynthesis pathway, derived from the shikimate pathway, is involved in plant defense mechanisms and physiological regulation. Up-regulated metabolites in T1 treatment were enriched in pathways including metabolic pathways, biosynthesis of secondary metabolites, and alkaloids derived from ornithine, lysine, and nicotinic acid. The activation of secondary metabolite pathways may reflect plant adaptation to environmental changes, while the regulation of alkaloid biosynthesis influences plant physiology (Kong et al., 2021). Alterations in the carbon metabolism pathway may affect energy metabolism and material circulation within the plant (Wu et al., 2024b). Up-regulated metabolites in T2 were enriched in general metabolic pathways. In T3 treatment, up-regulated metabolites were enriched in tryptophan metabolism and microbial metabolism. Changes in tryptophan metabolism could impact neurotransmitter synthesis and physiological regulation. The microbial metabolism pathways suggest interactions between microbial communities and the host organism (Lueders et al., 2006). These findings indicate that microbial metabolism plays a key role in T3 treatment, with the activation of secondary metabolite pathways reflecting adaptive responses to environmental conditions (Chamkhi et al., 2021).

The 16S and ITS sequencing data further elucidate how planting treatments influence microbial communities (Ge et al., 2023). T1, for example, led to an increase in genes related to transcription,

signal transduction, and stress-related processes, which suggests that soil microorganisms were actively responding to the altered rhizosphere environment (Korenblum et al., 2022; Li et al., 2023). Arbuscular mycorrhiza fungi (AMF) is a type of fungus that forms a symbiotic relationship with plant roots, promoting nutrient absorption and plant growth (Santander et al., 2017). The increased abundance of AMF in T1 highlights the role of these beneficial fungi in enhancing nutrient uptake and supporting plant growth through symbiotic relationships (Khan, 2020). This finding is consistent with other studies that emphasize the positive impact of AMF on plant nutrition and stress resilience in agricultural systems (Zhu et al., 2021). Saprophytic fungi, such as undefined saprophytes and wood-decaying fungi, play a key role in decomposing organic matter within ecosystems (Bollmann-Giolai et al., 2022). The increased abundance of these fungi in T1 may indicate more organic matter availability or changes in environmental conditions, fostering their growth (Mishra et al., 2024). The higher abundance of genes related to plant pathogens and endophytes may suggest a shift in the interactions between plants and microorganisms, possibly increasing the pressure from pathogens or altering the plant-endophyte relationship in T1 (Yue et al., 2023).

T2 treatment, which promoted genetic information processing and altered microbial gene replication, indicates an adaptive response to environmental conditions that may favor the proliferation of certain microbial groups (Li et al., 2021). The increased abundance of genes associated with animal pathogens and saprophytes in T2 might point to shifts in ecological niches and microbial competition, which could affect disease dynamics in the soil (Bai et al., 2022). These changes may reflect modifications in ecological niches, making certain fungal groups, such as wood-decaying fungi, more competitive in T2 (Pantigoso et al., 2022). The increased abundance of genes related to animal pathogens, saprophytes, and endophytes may indicate complex interactions between pathogens and fungal communities, or the influence of animal activities on soil microorganisms (Gouda et al., 2018).

T3 treatment revealed changes that favored the growth of dung saprotrophs and animal symbionts, which could be linked to the use of organic fertilizers or changes in soil pH and nutrient availability (Song et al., 2023). In T3, genes related to dung saprotrophs, undefined saprotrophs, and animal endosymbionts were more abundant. The higher abundance of fecal saprophytes and undefined saprotrophs could be linked to organic fertilizer application or increased animal activity under T3 (Qu et al., 2020). This suggests that the microbial community in T3 may have been more adapted to decomposing organic matter, which is a key process in maintaining soil structure and nutrient cycling (Zhou et al., 2023b). The increase in microbial pathways related to metabolism and nutrient acquisition further underscores the importance of microbial interactions in sustaining soil health and promoting plant growth (Yang et al., 2023).

The hierarchical microbial shifts revealed treatment-specific linkages between green manure inputs and functional microbiota assembly. T3's high-carbon amendment enriched Firmicutes and Fusarium, driving rapid organic matter decomposition, while T1's *V. myuros* favored acid-tolerant Acidobacteria and Glomeromycota, enhancing symbiotic nutrient acquisition. T2 uniquely selected Chloroflexi and Entrophospora, targeting recalcitrant lignin degradation. Bacterial communities exhibited stronger treatment divergence than fungi, where Ascomycota persistently dominated, suggesting bacteria better reflect short-term organic input changes. These contrasts highlight trade-offs: T3 prioritizes decomposition, T1/T2 enhance symbiosis, and crop-specific taxa selection guides precision soil management.

Implications for Sustainable Agriculture: The findings of this study align with the broader literature on the importance of crop rotation and intercropping for soil health (Kong et al., 2021). Intercropping with *V. myuros* and rotating with crops like rape offer clear benefits in terms of improving soil structure, reducing pest and disease incidence, and enhancing nutrient cycling (Yang et al., 2024). However, the excessive reliance on compound fertilizers, as highlighted in the study, poses risks of nutrient accumulation and soil degradation, which can impair long-term soil fertility (Liu et al., 2020). This underscores the need for balanced fertilizer use and careful management of organic and chemical inputs in tobacco farming to optimize both economic returns and environmental sustainability (Feng et al., 2023).

5 Conclusion

This study systematically investigated the effects of different tobacco cropping systems on soil rhizosphere ecology through integrated analysis of microbial communities and metabolite profiles. The results demonstrated that intercropping with *V. myuros* (T1) significantly enhanced symbiotic fungal associations (particularly Glomeromycota), while rotation with rape (T2) promoted the enrichment of lignin-degrading microbial consortia (Basidiomycota and Chloroflexi). The application of fermented cake fertilizer (T3) substantially increased microbial diversity and fostered cooperative metabolic networks, particularly in organic matter decomposition pathways. Metabolomic analysis revealed treatment-specific regulation of key metabolic pathways including phenylpropanoid biosynthesis, carbon metabolism, and amino acid metabolism, highlighting their crucial roles in nutrient cycling and plant-microbe interactions. These findings provide strong evidence that diversified cropping systems offer superior benefits over continuous monoculture, including improved soil health, enhanced nutrient availability, and better disease resistance while reducing dependence on synthetic fertilizers. The study suggests that optimal soil management strategies should be tailored to specific agricultural contexts, combining organic amendments with microbial conservation practices. Future research should focus on long-term field trials to validate these findings and develop practical guidelines for sustainable tobacco cultivation. This work contributes to our understanding of rhizosphere ecology and provides a scientific basis for developing more sustainable agricultural practices.

Data availability statement

The data that support the findings of this study are available on request from the corresponding author upon reasonable request. 16S-seq and ITS-seq raw data have been deposited in the NCBI sequence read archive (SRA) under project accession PRJNA1252381 and PRJNA1252388. The raw data for the metabolome of this study have been uploaded to MetaboLights (Number: MTBLS10471).

Author contributions

XL: Formal Analysis, Writing – original draft. YQ: Investigation, Writing – original draft. XW: Funding acquisition, Investigation, Writing – original draft. XC: Data curation, Funding acquisition, Investigation, Writing – original draft. FY: Investigation, Writing – original draft. HL: Data curation, Writing – original draft. CL: Data curation, Writing – original draft. YY: Resources, Writing – original draft. CY: Funding acquisition, Resources, Writing – review & editing. JY: Funding acquisition, Resources, Supervision, Writing – review & editing. PL: 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 National Natural Science Foundation of China (32000248), Key Science and Technology Project of China National Tobacco Corporation [110202201040 (XJ-11)], Talent Project of Guizhou Province (RCJD2020-7), and Bijie City Vegetable Industry Science and Technology Innovation Team. The funder China National Tobacco Corporation had no role in study design, data collection, analysis, interpretation, or the decision to publish the results.

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

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.2025.1595870/full#supplementary-material

References

Ahmed, M., Rauf, M., Mukhtar, Z., and Saeed, N. A. (2017). Excessive use of nitrogenous fertilizers: An unawareness causing serious threats to environment and human health. *Environ. Sci. Pollut. Res. Int.* 24, 26983–26987. doi: 10.1007/s11356-017-0589-7

Atiwesh, G., Parrish, C. C., Banoub, J., and Le, T.-A. T. (2022). Lignin degradation by microorganisms: A review. *Biotechnol. Prog.* 38, e3226. doi: 10.1002/btpr.3226

Babalola, O. O., Emmanuel, O. C., Adeleke, B. S., Odelade, K. A., Nwachukwu, B. C., Ayiti, O. E., et al. (2021). Rhizosphere microbiome cooperations: strategies for sustainable crop production. *Curr. Microbiol.* 78, 1069–1085. doi: 10.1007/s00284-021-02375-2

Bachy, C., Dolan, J. R., López-García, P., Deschamps, P., and Moreira, D. (2013). Accuracy of protist diversity assessments: Morphology compared with cloning and direct pyrosequencing of 18S rRNA genes and ITS regions using the conspicuous tintinnid ciliates as a case study. *ISME J.* 7, 244–255. doi: 10.1038/ismej.2012.106

Bai, B., Liu, W., Qiu, X., Zhang, J., Zhang, J., and Bai, Y. (2022). The root microbiome: Community assembly and its contributions to plant fitness. *J. Integr. Plant Biol.* 64, 230–243. doi: 10.1111/jipb.13226

Behnke, G. D., Kim, N., Zabaloy, M. C., Riggins, C. W., Rodriguez-Zas, S., and Villamil, M. B. (2021). Soil microbial indicators within rotations and tillage systems. *Microorganisms* 9, 1244. doi: 10.3390/microorganisms9061244

Bhattacharyya, P. N., and Jha, D. K. (2012). Plant growth-promoting rhizobacteria (PGPR): Emergence in agriculture. *World J. Microbiol. Biotechnol.* 28, 1327–1350. doi: 10.1007/s11274-011-0979-9

Bokulich, N. A., Joseph, C. M. L., Allen, G., Benson, A. K., and Mills, D. A. (2012). Next-generation sequencing reveals significant bacterial diversity of botrytized wine. *PloS One* 7, e36357. doi: 10.1371/journal.pone.0036357

Bokulich, N. A., Kaehler, B. D., Rideout, J. R., Dillon, M., Bolyen, E., Knight, R., et al. (2018). Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome* 6, 90. doi: 10.1186/s40168-018-0470-z

Bokulich, N. A., Subramanian, S., Faith, J. J., Gevers, D., Gordon, J. I., Knight, R., et al. (2013). Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nat. Methods* 10, 57–59. doi: 10.1038/nmeth.2276

Bollmann-Giolai, A., Malone, J. G., and Arora, S. (2022). Diversity, detection and exploitation: Linking soil fungi and plant disease. *Curr. Opin. Microbiol.* 70, 102199. doi: 10.1016/j.mib.2022.102199

Chamkhi, I., Benali, T., Aanniz, T., El Menyiy, N., Guaouguaou, F.-E., El Omari, N., et al. (2021). Plant-microbial interaction: The mechanism and the application of microbial elicitor induced secondary metabolites biosynthesis in medicinal plants. *Plant Physiol. Biochem.: PPB* 167, 269–295. doi: 10.1016/j.plaphy.2021.08.001

Chen, D., Zhou, Y., Wang, M., Mujtaba Munir, M. A., Lian, J., Yu, S., et al. (2022). Succession pattern in soil micro-ecology under tobacco (Nicotiana tabacum L.) continuous cropping circumstances in yunnan province of southwest China. *Front. Microbiol.* 12. doi: 10.3389/fmicb.2021.785110

Chen, H., Tang, X., Wang, T., Liao, W., Wu, Z., Wu, M., et al. (2024a). Calcium polypeptide mitigates Cd toxicity in rice via reducing oxidative stress and regulating pectin modification. *Plant Cell Rep.* 43, 163. doi: 10.1007/s00299-024-03253-4

Chen, P., Wang, L., Li, W.-H., Zhang, X.-X., Gao, H.-H., Zhou, X.-H., et al. (2024b). Micro-biological degradation and transformation of dissolved organic matter following continuous cropping of tobacco. *Front. Microbiol.* 15. doi: 10.3389/fmicb.2024.1319895

Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., and Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinf. (Oxford Engl.)* 27, 2194–2200. doi: 10.1093/bioinformatics/btr381

Fallah, N., Pang, Z., Lin, Z., Nyimbo, W. J., Lin, W., Mbuya, S. N., et al. (2023). Sustained organic amendments utilization enhances ratoon crop growth and soil quality by enriching beneficial metabolites and suppressing pathogenic bacteria. *Front. Plant Sci.* 14. doi: 10.3389/fpls.2023.1273546

Fang, Y., Zhang, L., Jiao, Y., Liao, J., Luo, L., Ji, S., et al. (2016). Tobacco rotated with rapeseed for soil-borne phytophthora pathogen biocontrol: mediated by rapeseed root exudates. *Front. Microbiol.* 7. doi: 10.3389/fmicb.2016.00894

Fatima, F., Pathak, N., and Rastogi Verma, S. (2014). An improved method for soil DNA extraction to study the microbial assortment within rhizospheric region. *Mol. Biol. Int.* 2014, 518960. doi: 10.1155/2014/518960

Feng, J., Chen, L., Xia, T., Ruan, Y., Sun, X., Wu, T., et al. (2023). Microbial fertilizer regulates C:N:P stoichiometry and alleviates phosphorus limitation in flue-cured tobacco planting soil. *Sci. Rep.* 13, 10276. doi: 10.1038/s41598-023-37438-w

Feng, Y., Chen, H., Fu, L., Yin, M., Wang, Z., Li, Y., et al. (2024). Green manuring enhances soil multifunctionality in tobacco field in southwest China. *Microorganisms* 12, 949. doi: 10.3390/microorganisms12050949

Ge, J., Li, D., Ding, J., Xiao, X., and Liang, Y. (2023). Microbial coexistence in the rhizosphere and the promotion of plant stress resistance: A review. *Environ. Res.* 222, 115298. doi: 10.1016/j.envres.2023.115298

Gouda, S., Kerry, R. G., Das, G., Paramithiotis, S., Shin, H.-S., and Patra, J. K. (2018). Revitalization of plant growth promoting rhizobacteria for sustainable development in agriculture. *Microbiol. Res.* 206, 131–140. doi: 10.1016/j.micres.2017.08.016

Hess, M., Sczyrba, A., Egan, R., Kim, T.-W., Chokhawala, H., Schroth, G., et al. (2011). Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. *Sci. (New York N.Y.)* 331, 463–467. doi: 10.1126/science.1200387

Holden, N. (2019). You are what you can find to eat: bacterial metabolism in the rhizosphere. *Curr. Issues Mol. Biol.* 30, 1–16. doi: 10.21775/cimb.030.001

Ilhan, Z. E., DiBaise, J. K., Isern, N. G., Hoyt, D. W., Marcus, A. K., Kang, D.-W., et al. (2017). Distinctive microbiomes and metabolites linked with weight loss after gastric bypass, but not gastric banding. *ISME J.* 11, 2047–2058. doi: 10.1038/ismej.2017.71

Jiang, Y., Zhang, R., Zhang, C., Su, J., Cong, W.-F., and Deng, X. (2022). Long-term organic fertilizer additions elevate soil extracellular enzyme activities and tobacco quality in a tobacco-maize rotation. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.973639

Kalam, S., Basu, A., Ahmad, I., Sayyed, R. Z., El-Enshasy, H. A., Dailin, D. J., et al. (2020). Recent understanding of soil acidobacteria and their ecological significance: A critical review. *Front. Microbiol.* 11. doi: 10.3389/fmicb.2020.580024

Khan, N. (2020). Plant-microbial interactions and their role in sustainable agriculture and sustainability of agriculture soils. *Recent Patents Food Nutr. Agric.* 11, 94–95. doi: 10.2174/2212798411022008061529 33

Kong, X., Xiao, Y., Zhi, Q., Liao, Y., He, B., Jian, H., et al. (2021). Soil fertility comparison between rotations of tobacco-rice (TR) and rape-rice (RR). *Agric. Sci.* 12, 1527–1537. doi: 10.4236/as.2021.1212097

Korenblum, E., Massalha, H., and Aharoni, A. (2022). Plant-microbe interactions in the rhizosphere via a circular metabolic economy. *Plant Cell* 34, 3168–3182. doi: 10.1093/plcell/koac163

Kumawat, K. C., Razdan, N., and Saharan, K. (2022). Rhizospheric microbiome: Bio-based emerging strategies for sustainable agriculture development and future perspectives. *Microbiol. Res.* 254, 126901. doi: 10.1016/j.micres.2021.126901

Lei, Y., Xiao, Y., Li, L., Jiang, C., Zu, C., Li, T., et al. (2017). Impact of tillage practices on soil bacterial diversity and composition under the tobacco-rice rotation in China. *J. Microbiol. (Seoul Korea)* 55, 349–356. doi: 10.1007/s12275-017-6242-9

Li, M., Guo, J., Ren, T., Luo, G., Shen, Q., Lu, J., et al. (2021). Crop rotation history constrains soil biodiversity and multifunctionality relationships. *Agricult. Ecosyst. Environ.* 319, 107550. doi: 10.1016/j.agee.2021.107550

Li, J.-H., Muhammad Aslam, M., Gao, Y.-Y., Dai, L., Hao, G.-F., Wei, Z., et al. (2023). Microbiome-mediated signal transduction within the plant holobiont. *Trends Microbiol.* 31, 616–628. doi: 10.1016/j.tim.2022.12.005

Lin, Y., Zhang, P., Wu, Q., Zhang, Y., Wei, Q., Sun, Y., et al. (2022). Leymus chinensis adapts to degraded soil environments by changing metabolic pathways and root exudate components. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.894346

Liu, L., Miao, Q., Guo, Y., Wang, C., Sun, J., Fan, Z., et al. (2024a). Bacterial and fungal communities regulated directly and indirectly by tobacco-rape rotation promote tobacco production. *Front. Microbiol.* 15. doi: 10.3389/fmicb.2024.1418090

Liu, S., Wang, J., Pu, S., Blagodatskaya, E., Kuzyakov, Y., and Razavi, B. S. (2020). Impact of manure on soil biochemical properties: A global synthesis. *Sci. Total Environ.* 745, 141003. doi: 10.1016/j.scitotenv.2020.141003

Liu, M., Xue, R., Wang, D., Hu, Y., Gu, K., Yang, L., et al. (2024b). Variations in different preceding crops on the soil environment, bacterial community richness and diversity of tobacco-planting soil. *Front. Microbiol.* 15. doi: 10.3389/fmicb.2024.1389751

Liu, M., Xue, R., Yang, C., Han, N., Hu, Y., Gu, K., et al. (2024c). Rotation with other crops slow down the fungal process in tobacco-growing soil. *Sci. Rep.* 14, 14160. doi: 10.1038/s41598-024-64735-9

Lu, J., Li, W., Yang, Y., Ye, F., Lu, H., Chen, X., et al. (2022). The impact of different rotation regime on the soil bacterial and fungal communities in an intensively managed agricultural region. *Arch. Microbiol.* 204, 142. doi: 10.1007/s00203-021-02615-w

Lueders, T., Kindler, R., Miltner, A., Friedrich, M. W., and Kaestner, M. (2006). Identification of bacterial micropredators distinctively active in a soil microbial food web. *Appl. Environ. Microbiol.* 72, 5342–5348. doi: 10.1128/AEM.00400-06

Magoč, T., and Salzberg, S. L. (2011). FLASH: Fast length adjustment of short reads to improve genome assemblies. *Bioinf. (Oxford Engl.)* 27, 2957–2963. doi: 10.1093/bioinformatics/btr507

McClure, R. S., Lee, J.-Y., Chowdhury, T. R., Bottos, E. M., White, R. A., Kim, Y.-M., et al. (2020). Integrated network modeling approach defines key metabolic responses of soil microbiomes to perturbations. *Sci. Rep.* 10, 10882. doi: 10.1038/s41598-020-67878-7

Mishra, S., Zhang, X., and Yang, X. (2024). Plant communication with rhizosphere microbes can be revealed by understanding microbial functional gene composition. *Microbiol. Res.* 284, 127726. doi: 10.1016/j.micres.2024.127726

Pantigoso, H. A., Newberger, D., and Vivanco, J. M. (2022). The rhizosphere microbiome: Plant-microbial interactions for resource acquisition. *J. Appl. Microbiol.* 133, 2864–2876. doi: 10.1111/jam.15686

Qiao, H., Jiao, B., Wang, J., Yang, Y., Yang, F., Geng, Z., et al. (2023). Comparative Analysis of miRNA Expression Profiles under Salt Stress in Wheat. *Genes* 14, 1586. doi: 10.3390/genes14081586

Qu, Q., Zhang, Z., Peijnenburg, W. J. G. M., Liu, W., Lu, T., Hu, B., et al. (2020). Rhizosphere microbiome assembly and its impact on plant growth. *J. Agric. Food Chem.* 68, 5024–5038. doi: 10.1021/acs.jafc.0c00073

Rampadarath, S., Bandhoa, K., Puchooa, D., Jeewon, R., and Bal, S. (2018). Metatranscriptomics analysis of mangroves habitats around Mauritius. *World J. Microbiol. Biotechnol.* 34, 59. doi: 10.1007/s11274-018-2442-7

Santander, C., Aroca, R., Ruiz-Lozano, J. M., Olave, J., Cartes, P., Borie, F., et al. (2017). Arbuscular mycorrhiza effects on plant performance under osmotic stress. *Mycorrhiza* 27, 639–657. doi: 10.1007/s00572-017-0784-x

Sawers, R. J. H., Ramírez-Flores, M. R., Olalde-Portugal, V., and Paszkowski, U. (2018). The impact of domestication and crop improvement on arbuscular mycorrhizal symbiosis in cereals: Insights from genetics and genomics. *New Phytol.* 220, 1135–1140. doi: 10.1111/nph.15152

Shi, M., Hao, S., Wang, Y., Zhang, S., Cui, G., Zhang, B., et al. (2024). Plant growth-promoting fungi improve tobacco yield and chemical components by reassembling rhizosphere fungal microbiome and recruiting probiotic taxa. *Environ. Microb.* 19, 83. doi: 10.1186/s40793-024-00629-7

Song, Y., Sun, L., Wang, H., Zhang, S., Fan, K., Mao, Y., et al. (2023). Enzymatic fermentation of rapeseed cake significantly improved the soil environment of tea rhizosphere. *BMC Microbiol.* 23, 250. doi: 10.1186/s12866-023-02995-7

Su, Y., Zi, H., Wei, X., Hu, B., Deng, X., Chen, Y., et al. (2022). Application of manure rather than plant-origin organic fertilizers alters the fungal community in continuous cropping tobacco soil. *Front. Microbiol.* 13. doi: 10.3389/fmicb.2022.818956

Sun, L., Tsujii, Y., Xu, T., Han, M., Li, R., Han, Y., et al. (2023a). Species of fast bulk-soil nutrient cycling have lower rhizosphere effects: A nutrient spectrum of rhizosphere effects. *Ecology* 104, e3981. doi: 10.1002/ecy.3981

Sun, L., Wang, S., Narsing Rao, M. P., Shi, Y., Lian, Z.-H., Jin, P.-J., et al. (2023b). The shift of soil microbial community induced by cropping sequence affect soil properties and crop yield. *Front. Microbiol.* 14. doi: 10.3389/fmicb.2023.1095688

Tiemann, L. K., Grandy, A. S., Atkinson, E. E., Marin-Spiotta, E., and McDaniel, M. D. (2015). Crop rotational diversity enhances belowground communities and functions in an agroecosystem. *Ecol. Lett.* 18, 761–771. doi: 10.1111/ele.12453

Wang, S., Cheng, J., Li, T., and Liao, Y. (2020). Response of soil fungal communities to continuous cropping of flue-cured tobacco. *Sci. Rep.* 10, 19911. doi: 10.1038/s41598-020-77044-8

Wang, X., Li, Z., Jin, H., Huang, X., Xiao, X., Zi, S., et al. (2023). Effects of mushroom-tobacco rotation on microbial community structure in continuous cropping tobacco soil. *J. Appl. Microbiol.* 8, lxad088. doi: 10.1093/jambio/lxad088

Wang, G., Sheng, L., Zhao, D., Sheng, J., Wang, X., and Liao, H. (2016). Allocation of nitrogen and carbon is regulated by nodulation and mycorrhizal networks in soybean/maize intercropping system. *Front. Plant Sci.* 7, 1901. doi: 10.3389/fpls.2016.01901

Want, E. J., Masson, P., Michopoulos, F., Wilson, I. D., Theodoridis, G., Plumb, R. S., et al. (2013). Global metabolic profiling of animal and human tissues via UPLC-MS. *Nat. Protoc.* 8, 17–32. doi: 10.1038/nprot.2012.135

Wu, H., Cui, H., Fu, C., Li, R., Qi, F., Liu, Z., et al. (2024a). Unveiling the crucial role of soil microorganisms in carbon cycling: A review. *Sci. Total Environ.* 909, 168627. doi: 10.1016/j.scitotenv.2023.168627

Wu, G., Hou, Q., Zhan, M., Zhang, H., Lv, X., and Xu, Y. (2024b). Metabolome regulation and restoration mechanism of different varieties of rice (Oryza sativa L.) after lindane stress. *Sci. Total Environ.* 913, 169839. doi: 10.1016/j.scitotenv.2023.169839

Xiong, Z., Zhu, D., Lu, Y., Lu, J., Liao, Y., Ren, T., et al. (2023). Continuous potassium fertilization combined with straw return increased soil potassium availability and risk of potassium loss in rice-upland rotation systems. *Chemosphere* 344, 140390. doi: 10.1016/j.chemosphere.2023.140390

Yan, S., Niu, Z., Yan, H., Zhang, A., and Liu, G. (2019). Transcriptome sequencing reveals the effect of biochar improvement on the development of tobacco plants before and after topping. *PloS One* 14, e0224556. doi: 10.1371/journal.pone.0224556

Yang, S., Liu, H., Xie, P., Wen, T., Shen, Q., and Yuan, J. (2023). Emerging pathways for engineering the rhizosphere microbiome for optimal plant health. *J. Agric. Food Chem.* 71, 4441–4449. doi: 10.1021/acs.jafc.2c08758

Yang, L., Wang, C., He, X., Liang, H., Wu, Q., Sun, X., et al. (2024). Multi-year crop rotation and quicklime application promote stable peanut yield and high nutrient-use efficiency by regulating soil nutrient availability and bacterial/fungal community. *Front. Microbiol.* 15. doi: 10.3389/fmicb.2024.1367184

Yue, H., Yue, W., Jiao, S., Kim, H., Lee, Y.-H., Wei, G., et al. (2023). Plant domestication shapes rhizosphere microbiome assembly and metabolic functions. *Microbiome* 11, 70. doi: 10.1186/s40168-023-01513-1

Zhang, S., Li, S., Meng, L., Liu, X., Zhang, Y., Zhao, S., et al. (2024). Root exudation under maize/soybean intercropping system mediates the arbuscular mycorrhizal fungi diversity and improves the plant growth. *Front. Plant Sci.* 15. doi: 10.3389/fpls.2024.1375194

Zhang, X., Jiang, H., Wan, X., and Li, Y. (2020). The effects of different types of mulch on soil properties and tea production and quality. *J. Sci. Food Agric.* 100, 5292–5300. doi: 10.1002/jsfa.10580

Zhou, Q., Wang, L. C., Xing, Y., Ma, S. M., Zhang, X. D., Chen, J., et al. (2018). Effects of intercropping Chinese milk vetch on functional characteristics of soil microbial community in rape rhizosphere. *Ying Yong Sheng Tai Xue Bao = J. Appl. Ecol.* 29, 909–914. doi: 10.13287/j.1001-9332.201803.031

Zhou, Y., Yang, Z., Liu, J., Li, X., Wang, X., Dai, C., et al. (2023a). Crop rotation and native microbiome inoculation restore soil capacity to suppress a root disease. *Nat. Commun.* 14, 8126. doi: 10.1038/s41467-023-43926-4

Zhou, H., Zhang, M., Yang, J., Wang, J., Chen, Y., and Ye, X. (2023b). Returning ryegrass to continuous cropping soil improves soil nutrients and soil microbiome, producing good-quality flue-cured tobacco. *Front. Microbiol.* 14. doi: 10.3389/fmicb.2023.1257924

Zhu, M., Ding, Y., Li, X., Xiao, Y., Zhao, Z., and Li, T. (2021). Biodiversity of root endophytic fungi from oxyria sinensis grown in metal-polluted and unpolluted soils in yunnan province, Southwestern China. *Plants (Basel Switzerland)* 10, 2731. doi: 10.3390/plants10122731

Zhu, B., Gu, H., He, J., Li, F., Yu, J., Liu, W., et al. (2023). The impact of smash-ridge tillage on agronomic traits of tobacco plants, soil enzymatic activity, microbial community structure, and functional diversity. *Plant Signaling Behav.* 18, 2260640. doi: 10.1080/15592324.2023.2260640

Zhuang, Y., Wang, H., Tan, F., Wu, B., Liu, L., Qin, H., et al. (2024). Rhizosphere metabolic cross-talk from plant-soil-microbe tapping into agricultural sustainability: Current advance and perspectives. *Plant Physiol. Biochem.: PPB* 210, 108619. doi: 10.1016/j.plaphy.2024.108619

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 30 June 2025
DOI 10.3389/fpls.2025.1588485

Check for updates

OPEN ACCESS

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

REVIEWED BY
Shu-Yun Chen,
National Chung Hsing University, Taiwan
Arvind Mohanan,
Sri Karan Narendra Agriculture University,
India

*CORRESPONDENCE
Somashekhar M. Punnuri
punnuris@fvsu.edu

RECEIVED 06 March 2025
ACCEPTED 09 May 2025
PUBLISHED 30 June 2025

CITATION
Vutla S, Knoll JE, Sankuratri A, Nayak RG, Liu L, Chee PW, Karthikeyan R, Fakrudin B, Thudi M, Timper P, Harris-Shultz K, Wallace JG, Singh H, Singh B and Punnuri SM (2025) Mapping QTLs for *Pyricularia* leaf spot, nematode resistance, and yield related traits in pearl millet [*Cenchrus americanus* (L.) Morrone]. *Front. Plant Sci.* 16:1588485. doi: 10.3389/fpls.2025.1588485

COPYRIGHT
© 2025 Vutla, Knoll, Sankuratri, Nayak, Liu, Chee, Karthikeyan, Fakrudin, Thudi, Timper, Harris-Shultz, Wallace, Singh, Singh and Punnuri. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Mapping QTLs for *Pyricularia* leaf spot, nematode resistance, and yield related traits in pearl millet [*Cenchrus americanus* (L.) Morrone]

Sairam Vutla[1], Joseph E. Knoll[2], Anvesh Sankuratri[1], Risha G. Nayak[1], Limei Liu[3,4], Peng W. Chee[3], Raghupathy Karthikeyan[5], Bashasab Fakrudin[6], Mahendar Thudi[1], Patricia Timper[2], Karen Harris-Shultz[2], Jason G. Wallace[7,8], Hari Singh[1], Bharat Singh[1] and Somashekhar M. Punnuri[1]*

[1]College of Agriculture, Family Sciences and Technology, 1005 State University Dr. Fort Valley State University, Fort Valley, GA, United States, [2]Crop Genetics and Breeding Research Unit, United States Department of Agriculture (USDA)-Agricultural Research Service, Tifton, GA, United States, [3]Institute of Plant Breeding, Genetics, and Genomics, University of Georgia, Tifton, GA, United States, [4]School of Plant and Environmental Sciences, Virginia Polytechnic Institute and State University, Blacksburg, VA, United States, [5]Department of Agricultural Sciences, College of Agriculture, Forestry and Life Sciences, Clemson University, Clemson, SC, United States, [6]University of Horticultural Sciences, Bagalkot, Karnataka, India, [7]Department of Crop & Soil Sciences, University of Georgia, Athens, GA, United States, [8]Center For Applied Genetic Technologies (CAGT), University of Georgia, Athens, GA, United States

Pearl millet [*Cenchrus americanus* (L.) Morrone, formerly *Pennisetum glaucum* (L.) R. Br.] is the sixth most important cereal globally and is used for forage and feed in the U.S. To identify genomic regions governing important physiological, agronomic and yield related traits, a recombinant inbred line population derived from the cross between Tift $99D_2B_1$ × Tift 454 was phenotyped in the field in 2006, 2007 and 2013. In addition, the population was phenotyped for root-knot nematode resistance in the greenhouse during 2009. Using a previously generated genetic map containing 505 single nucleotide polymorphism markers and composite interval mapping, we identified 45 QTLs for eight traits (plant height, stem diameter, days to heading, panicle diameter, panicle length, 1000 seed weight, *Pyricularia* leaf spot disease, and root-knot nematode egg mass) across almost all linkage groups. These QTLs explained 6.31 to 32.51% of phenotypic variance for each trait and were consistently detected over different environments. Plant height and days to heading were colocalized on LG2 and LG5 showing maturity and plant height are linked and influence each other, similarly to other cereal crops. Interestingly, 5 of 19 QTLs linked to plant height, stem diameter, panicle diameter, and panicle length colocalized to the same locations on LG3, indicating breeding for one trait could simultaneously improve the other. The markers and genes identified in the present study can be used in developing high yielding pearl millet varieties using marker-assisted selection.

KEYWORDS

pearl millet, QTLs, physiological traits, nematode resistance, yield-related traits

Introduction

Pearl millet [*Cenchrus americanus* (L.) Morrone, formerly *Pennisetum glaucum* (L.) R. Br.; $2n = 2x = 14$] is the sixth most important cereal globally, covering approximately 30 million hectares in Asia and Africa (Satyavathi et al., 2021). Globally, it serves as a staple food grain and fodder for over 90 million people (Srivastava et al., 2020; Satyavathi et al., 2021). It is grown in the United States, primarily for grazing, hay, cover crops, and wildlife. Pearl millet grain has a high nutritional value, characterized by high metabolizable energy, protein, micronutrient content and a low glycemic index (Agarwal et al., 2023). Nutritionally, the grain is gluten-free, making it suitable for specific dietary needs (Rai et al., 2008; Pei et al., 2022). On average, pearl millet contains approximately 12% protein, 69% carbohydrates, 5% lipids, 2.5% fiber, and 2.5% ash in terms of proximate nutritional composition (Andrews and Kumar, 1992; Tomar et al., 2021). Pearl millet has remarkable resilience to drought and heat stress, enabling cultivation in some of the most challenging environments (Gangashetty et al., 2016; Varshney et al., 2017). The above-ground portion of pearl millet is used as forage for dairy cows and other classes of livestock (Hancock et al., 2018). Pearl millet grain can also be used as feed for wildlife, poultry, beef cattle, and pigs (Andrews and Kumar, 1992). It has been grown in the southern U.S., particularly in Georgia, where there is potential for use by the poultry industry (Davis et al., 2003). However, there are few commercial grain hybrids available, (a notable exception is TifGrain 102), indicating a scope to develop hybrids suitable for this region (Hanna et al., 1991).

Pearl millet can grow on marginal soils with low fertility and organic matter content (Andrews and Kumar, 1992). Despite its positive attributes, pearl millet grain yields remain low, averaging around 0.85 tons/ha, which can be attributed to abiotic stresses (drought, low soil fertility) and biotic stresses such as nematodes, Striga (*Striga hermonthica*), millet head miner (*Heliocheilus albipunctella*), downy mildew (*Sclerospora graminicola*), leaf spot (*Pyricularia grisea*) and other diseases (FAOSTAT, 2020). By selecting and breeding pearl millet varieties with optimal panicle length, panicle diameter, and high 1000 seed weight, it is possible to enhance the resilience and productivity of this crop. These targeted improvements in yield components can lead to significant gains in grain yield, even under the challenging conditions posed by biotic and abiotic stresses (Poncet et al., 2000; Yadav et al., 2002, 2004; Vengadessan et al., 2013). Among biotic stresses in the southern region of the United States, plant-parasitic root-knot nematodes (RKN), *Meloidogyne incognita*, are important constraints in cropping systems (Timper and Wilson, 2006). Earlier efforts were made to identify root-knot nematode-resistant germplasm lines, and as a result, the inbred lines Tift 454 and Tift 99B were reported to be highly resistant and susceptible to RKN, respectively (Hanna et al., 2005a). In addition to root-knot nematodes, *Pyricularia* leaf spot (*Pyricularia grisea*) represents a yield limiting factor in pearl millet production (Timper et al., 2002). The commercial grain hybrid 'TifGrain 102' has been reported to exhibit strong resistance to *Pyricularia* leaf spot, providing a valuable resource for managing this disease (Hanna and Wells, 1989; Punnuri et al., 2016).

Understanding the underlying genetics and identifying the key genomic regions responsible for agronomic and yield-related traits will accelerate the development of climate-resilient varieties and enhance genetic gain (Varshney et al., 2018; Thudi et al., 2021; Yadav et al., 2021). Until recently, pearl millet was considered an orphan crop as it lagged behind sorghum (*Sorghum bicolor* (L.) Moench.) and other major cereals in terms of genetic improvement due to a lack of genomic resources (Passot et al., 2016). However, in recent years, genomic resources have dramatically expanded for this crop. High-density genetic maps for pearl millet were created using genotyping-by-sequencing (GBS) markers (Azhaguvel, 2001; Moumouni et al., 2015; Punnuri et al., 2016). Furthermore, significant efforts have been made to map the genomic regions responsible for abiotic stresses (Yadav et al., 2002; Sharma et al., 2011; Sehgal et al., 2012), and biotic stresses (Ambawat et al., 2016; Punnuri et al., 2016; Chelpuri et al., 2019). The genes involved in reducing internode length influencing plant height have been identified by map-based cloning and comparative mapping (Parvathaneni et al., 2013). In addition, with the availability of the pearl millet genome sequence and resequencing information of ~1000 germplasm lines, including wild species, pearl millet is now considered a genomics resource-rich crop (Varshney et al., 2017; Burgarella et al., 2018). These significant advances in genome sequencing in addition to high-density genetic maps will support genomics-assisted breeding (Punnuri et al., 2024).

There are very few studies that have mapped genomic regions for agronomic and yield-related traits using bi-parental populations arising from two dwarf, early maturing grain-type parents, particularly those used in producing a commercial hybrid. TifGrain 102 is a commercialized hybrid developed from Tift $99D_2B_1$ and Tift 454, used in the current study, and has shown promising results in terms of its uniformity and yield stability. Therefore, mapping quantitative trait loci (QTLs) associated with yield and its components in this recombinant inbred line (RIL) population will provide greater insight into useful alleles and traits associated with these two parents. In this study, we report the QTLs for key agronomic and yield-related traits, as well as root-knot nematode resistance and *Pyricularia* leaf spot resistance, using previously generated GBS data (Punnuri et al., 2016) and phenotyping data generated during 2006, 2007, 2009, and 2013. In addition, we also report potential candidate genes in these QTL regions. The markers and genes reported in the present study can be used for genomics-assisted breeding of pearl millet for these important traits.

Materials and methods

Plant material

A RIL population comprising of 225 lines derived from the parental genotypes Tift $99D_2B_1$ (female) and Tift 454 (Punnuri et al., 2016), was used for mapping QTLs for agronomic, yield related traits and biotic stress resistance. In brief, both parents are

dwarf, early maturing grain-types that have been successfully used in producing a commercial hybrid known as TifGrain 102, where Tift 99B serves as the male-fertile maintainer line for the male-sterile Tift 99A (Hanna et al., 2005a, 2005b). Tift 454 is highly resistant to root-knot nematode (*Meloidogyne incognita*) (Hanna et al., 2005a), but it is susceptible to *Pyricularia* leaf spot. In contrast, Tift 99B is resistant to rust (*Puccinia substriata* var. *indica*) and other diseases, including leaf spot caused by *Pyricularia grisea*, but is vulnerable to root-knot nematode. Although both Tift 99B and Tift 454 differ only slightly in various traits, their differences fall within the typical range observed among cultivated grain-type pearl millet inbreds in the United States. Additionally, TifGrain 102 has shown strong resistance to *Pyricularia* leaf spot, making it a valuable resource for improving disease resistance in breeding programs (Hanna and Wells, 1989; Punnuri et al., 2016). F_6 and F_7 lines were used for phenotyping traits.

Experimental design and trait phenotyping

A total of seven yield-related traits were phenotyped in the field over three years, though not all traits were assessed each year. In addition, the RIL population was also phenotyped for root-knot nematode resistance in the greenhouse in 2009 and for *Pyricularia* leaf spot disease incidence in 2013 (Supplementary Table S1). All experiments were arranged as randomized complete block designs. During 2006 and 2007 the F_6 and F_7 RIL populations, respectively were evaluated at Tifton (31.4505° N, 83.5085° W), Georgia, USA, with two replications each year. In 2013, 179-F_7 RILs were evaluated at Old Farm (32.5344° N, 83.8961° W), Fort Valley State University (FVSU), Fort Valley, Georgia, with three replications.

In 2006, days to heading, grain yield (g/head), and plant height (cm) were collected from 225 F_6 RILs, Tift $99D_2B_1$, Tift 454, HGM100, and TifGrain102. Plant height was measured from the ground to the tip of the panicle. All three traits were measured on ten representative plants per plot and the average was taken. In 2007, the same population was planted as in the year 2006 but only 222 F_7 lines were phenotyped due to poor germination and plant stand count (Supplementary Table S2). Phenotypic data collected in 2007 included grain yield (g/head), panicle diameter (mm), and panicle length (cm). In 2013, only 179 RILs from 184 RILs planted were phenotyped due to poor seed germination and plant losses caused by unfavorable field conditions. The traits evaluated in 2013 were plant height (cm), panicle length (cm), stem diameter (mm), panicle diameter (mm), days to heading, and seed weight for 1000 seeds (g). Plant height was measured from the base of the plant above ground till the flag leaf base (peduncle of the panicle). Panicle diameter and stem diameter were measured using digital Vernier calipers. Days to heading was counted from sowing till the complete emergence of panicle in 10 plants of each plot. Weight of 1000 seed was measured on one composite seed sample from each plot. Seeds were counted using a digital seed counter (Old Mill Equipment Co, USA).

Evaluation for *Pyricularia* leaf spot disease

A F_7 population (N=179) was grown in the field at Fort Valley (32°31'09.9"N 83°52'02.1"W) in 2013 in a randomized complete block design with three replications and was used to assess leaf spot disease. The same experimental block that was used for evaluating above traits in 2013 was also used for assessing *Pyricularia* leaf spot incidence. *Pyricularia* leaf spot infection occurred under natural conditions because of the rainy, humid weather during our experiment. Ten plants in each plot were visually scored and an average rating for the plot was used for analysis. The disease symptoms were very distinct on all RILs and the parental line (Punnuri et al., 2016). The disease scoring was performed as described by the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) using a 1–9 scale (Thakur et al., 2011), where 1 indicates no symptoms and 9 indicates complete plant death from disease. As disease progress is affected by growth stage, some of the late-maturing lines showed less severe symptoms. Therefore, disease scores for later-maturing RILs were adjusted based on their maturity and disease progress curve (Wilson and Hanna, 1992).

Evaluation for nematode resistance

In 2009, a set of 180 RILs along with both inbred parents and their F_1 hybrid were planted in 10 cm diameter pots in the greenhouse in a randomized complete block design with three replications. The susceptible parent, Tift $99D_2B_1$, was planted in three pots per replication, while the other lines were planted one pot per replication. Soil sterilization and greenhouse management was as described (Timper and Wilson, 2006). One vigorous plant per pot was retained after 10 days. *M. incognita* race 3, was cultured on eggplant (*Solanum melongena*) cv. Florida Market and 8000 eggs (4000 were added on the first day and another 4000 was added after two days) per pot were inoculated. Plants were removed from pots eight weeks after inoculation (two nematode generations) and examined for egg masses on roots. For this, roots were rinsed in water and stained in a 0.05% phloxine B solution for 3-5 minutes during which egg masses are stained bright red. Egg mass was estimated on a 0 – 5 scale (Holbrook et al., 1983) as follows: 0 = no egg mass, 1 = 1 – 2 egg masses, 2 = 3 – 10 egg masses, 3 = 11 – 30 egg masses, 4 = 31 – 100 egg masses, and 5 = greater than 100 egg masses. After rating the trials from first two replications, we observed that the egg masses were very small, so we waited two more weeks to let them grow in replication 3. The data recorded from each trial (1, 2 and 3) was averaged and used for analysis.

Statistical analysis for field and greenhouse data

Mean values and ranges were calculated among RILs for all traits for the respective years. Correlations between traits were

calculated with in each year using the CORR procedure of SAS v.9.4 (SAS Institute, Inc.).

The VARCOMP procedure of SAS was used to estimate variance components with the default method (MIVQUE0) and with RILs and replications as random factors. Variance components were used to calculate broad sense heritability (H^2) (Equation 1) using the following formula:

$$H^2 = \sigma_g^2/(\sigma_g^2 \times \sigma_\varepsilon^2/r) \quad (1)$$

where σ_g^2 is the variance due to genotype (RILs) and σ_ε^2 is the error variance and r is the number of replications.

The GLM procedure of SAS was used for the analysis of variance (Two-way ANOVA with a 5% level of significance) with genotype and replication as fixed effects. Fisher's LSD test was done to determine differences among entries and replications with a 5% level of significance.

G-model based single marker analysis

The genotype and phenotype dataset were also utilized for marker-trait analysis using the G-Model (Bernardo, 2013). SNP data were recoded as '1', '-1', '0', and '0.5' for the Tift 454 allele, the Tift 99D_2B_1 allele, a heterozygous allele, and missing data, respectively. The significance level for the marker effect was 0.00001 as recommended (Bernardo, 2013). RIL phenotype data that was not normally distributed was transformed by using $\log_{10}$ (x + 1).

Genetic map construction and QTL analysis

The GBS-based SNP data generated by Punnuri et al. (2016) on the RIL population identified 1,191 core markers. To avoid problems created by segregation distortion, we excluded all markers with segregation distortion while constructing the genetic linkage map using a chi-square test with alpha of < 0.05 and hence used 505 non-distorted markers from the 1,191 core markers previously deployed. Each linkage group (LG) was grouped as per 7 chromosomes named in Punnuri et al. (2016). QTL analysis was carried out by composite interval mapping in Windows QTL Cartographer software v2.5_011 (Wang et al., 2012). For this, the SNP calls were encoded as '2', '0', and '-' for the Tift 454 allele, the Tift 99D_2B_1 allele, and missing data, respectively. Missing phenotypes were represented by "period". The LOD ($\log_{10}$ likelihood ratio) thresholds of significant QTLs were determined by performing 1000 permutations using a Type I error set at $p < 0.001$. The percentage of phenotypic variance explained by a QTL was calculated by multiplying the R^2 by 100. QTLs explaining more than 10% phenotypic variation explained (PVE) were considered major QTLs and < 10% phenotypic variation was considered as minor QTLs. The linkage map showing the identified QTLs was constructed using MapChart v2.32 (Voorrips, 2002). We chose to use a genetic map instead of a physical map in case the parental lines showed any significant deviations from the reference genome (insertions, deletions, inversions, etc.).

Candidate gene analysis

The pearl millet genome sequence database available publicly (NCBI database) was used for gene identification. An in-house script was developed to identify candidate genes that resided near the identified QTLs. Genes that were within the QTL were identified and compiled for their primary functional annotation. Only genes with known functional annotations and Gene Ontology (GO) terms are reported.

Results

Performance of the population, variance components and broad-sense heritability

We identified significant variation among parents and RILs for most of the traits studied (Table 1). Tift 454 had higher average values than Tift 99B for all traits measured except stem diameter (2013), days to heading (2006), grain yield (2006), and nematode egg mass. For nematode egg mass, Tift 99B had a mean of 1.89 (out of 5), while Tift 454 had a mean of 0.0. For all agronomic traits, the F_1 hybrid plants had greater values than either of its parents except for days to heading (2006 and 2013), *Pyricularia* leaf spot, and RKN egg mass. The mean values for most traits among the RILs were intermediate between the two parents. We observed significant variation in plant height between the 2006 and 2013 phenotypic evaluations due to differences in measurement. During the 2006 field evaluations at Tifton, plant height among RILs varied between 35-155 cm with an average height of 86.1 cm. In the 2013 field evaluations at Fort Valley, the height among RILs varied between 42.5-109 cm with an average height of 73.4 cm. The broad-sense heritability of the plant height trait at both locations was high (H^2 = 0.877 in 2016 and H^2 = 0.756 in 2013). The stem diameter measured in 2013 varied between 1.50 to 6.74 mm with an average of 4.29 mm and had a heritability of 0.837 (Table 1). During 2013, days to heading varied between 42.7 to 57.7 days with an average of 47 days, while days to heading at Tifton in 2006 varied between 39.1 to 59.5 days with an average of 45.6 days. Nevertheless, heritability for days to heading was slightly higher in 2006 (H^2 = 0.853). We observed heritability of 0.809 and 0.852 during 2013 and 2007, respectively, for panicle diameter. During 2013, 1000 seed weight for RILs varied between 5.9 to 12.2 g with an average weight of 8.7 g and heritability of 0.829. Grain yield per panicle varied between 0.14 to 15.3 g with an average of 3.08 g during 2007, while in 2006, grain yield varied between 0.07 to 12.5 g per head with an average of 3.83 g. The heritability for grain yield was 0.789 and 0.824 during 2007 and 2006, respectively (Table 1). The egg mass rating values of RILs showed a range of 0 to 3.33 with a mean of 1.37 for all trials and heritability of 0.718.

Genetic correlations between different traits

Correlations between traits were observed within each year (Supplementary Table S3). The highest correlations in 2013 were observed between stem diameter and panicle diameter (0.59), plant

10.3389/fpls.2025.1588485

TABLE 1 Summary of phenotypic variation among parents, F_1 and RILs. Mean values, range, standard deviation and heritability values are for the RILs.

Trait	Year	Tift99B	Tift454	F_1	RILs				
		Mean	Mean	Mean	Minimum	Maximum	Mean	St. Dev.	H^2
Plant Height (cm)	2006	70.0	110	125	35.0	155	86.1	16.0	0.877
	2013	63.2	80.5	108	42.5	109	73.4	12.3	0.756
Stem Diameter (mm)	2013	4.41	3.58	6.61	1.50	6.74	4.29	1.14	0.837
Days To Heading (d)	2006	45.1	43.2	44.6	39.1	59.5	45.6	4.77	0.853
	2013	44.0	48.3	47.0	42.7	57.7	47.0	3.31	0.766
Panicle Length (cm)	2007	18.5	26.8	29.8	14.6	30.4	22.2	3.29	0.859
	2013	13.3	21.8	26.2	11.5	26.7	17.6	3.20	0.678
Panicle Diameter (mm)	2007	23.5	24.4	27.5	16.7	31.4	22.2	2.69	0.852
	2013	12.7	15.8	18.0	2.76	25.2	14.4	3.62	0.809
Grain Yield (g/head)	2006	5.06	4.82	7.09	0.07	12.5	3.83	2.26	0.824
	2007	2.88	5.96	9.53	0.14	15.3	3.08	2.14	0.789
1000 Seed Weight (g)	2013	8.83	10.33	12.7	5.9	12.2	8.7	0.9	0.829
Pyricularia Leaf Spot (1-9)	2013	3.0	7.3	2.0	1.0	9.0	5.6	1.8	0.939
Egg mass Avg_trial (1,2,3)	2009	1.89	0.0	1.33	0.0	3.33	1.37	1.05	0.718
Egg mass trial (1&2)	2009	1.66	0.0	1.50	0.0	3.50	1.25	1.15	0.731
Egg mass trial 3	2009	2.33	0.0	1.00	0.0	4.0	1.60	1.40	–

H^2, Broad-sense heritability; St. Dev., Standard Deviation.

height and stem diameter (0.49), and plant height and panicle length (0.56). In 2007, moderate correlations were observed between panicle length and panicle diameter (0.42), as well as grain yield per head and panicle diameter (0.40). In 2006, all traits measured were weakly correlated.

G-model results

Single marker analysis using the G model revealed several markers associated with the different traits measured in the pearl millet mapping population (Table 2). There were five markers directly associated with five different traits on four linkage groups. SNP marker S3_2921 was associated with plant height (2013), panicle length (2013), and panicle diameter (2007). SNP marker S3_2737 was associated with panicle length (2007), panicle diameter (2013), and stem diameter (2013). The markers associated with panicle length and panicle diameter for 2007 and 2013 and plant height (2013) on linkage group 3 were in an overlapping region.

Genetic map construction and QTLs mapping

A total of 505 non-distorted markers were mapped on the seven linkage groups of pearl millet. The number of SNP markers mapped per linkage group (LG) ranged from 35 to 89 with an average of 72.14 markers per linkage group (Supplementary Figure S1; Supplementary Table S4). LG3 had the most markers (89 SNPs), followed by LG7 (85 SNPs) and LG5 (82 SNPs). Exclusion of segregation distortion markers reduced our original map to very few markers on all linkage groups. Regions such as LG3 and LG4 were disproportionately affected and LG4 also had less polymorphic markers to begin with due to shared ancestry. We identified QTLs on all chromosomes (LGs) except chromosome 4 for all traits studied across different years: plant height (2006 & 2013) on LG2, LG3, and LG5; stem diameter (2013) on LG3 and LG7; days to heading (2006 & 2013) on LG2 and LG5; panicle diameter (2007 & 2013) on LG1, LG2 and LG3; panicle length (2007 & 2013) on LG3 and LG7; grain yield (2006 & 2007) on LG5 and LG6; 1000 seed weight (2013) on LG7; *Pyricularia* leaf spot disease (2013) on LG5; and nematode egg mass (2009) on LG2 (Table 3). In total, we identified 45 QTLs on six linkage groups using composite interval mapping with LOD> 3.00 (Figure 1).

QTLs for morpho-physiological traits: plant height and stem diameter

We identified a total of seven QTLs for plant height on linkage groups (LG) 2, 3 and 5 (Table 3). The phenotypic variance explained by these QTLs ranged from 6.38-10.16%. Further,

10.3389/fpls.2025.1588485

TABLE 2 Single nucleotide polymorphism markers identified as being associated with pearl millet traits using the G-Model.

Trait	Year	Associated SNP	LG	Position (cM)	Marker effect	p-value
Plant Height	2006	S2_6957	2	39.53	0.63	$9x10^{-7}$
	2013	S3_2921	3	14.79	4.27	$5x10^{-7}$
Stem Diameter	2013	S3_2737	3	14.36	0.4	$7.8x10^{-6}$
Days To Heading	2006	S5_0677	5	3.6	-0.02	$8x10^{-7}$
Panicle Diameter	2013	S3_2737	3	14.36	1.76	0
	2007	S3_2921	3	14.79	1.49	0
Panicle Length	2007	S7_0716	7	14.15	-0.81	$3.5x10^{-6}$
	2007	S3_2737	3	14.36	1.63	0
	2013	S3_2921	3	14.79	1.48	0

among these QTLs, the QTL based on 2013 phenotyping data on LG5 explained the maximum phenotypic variation, 10.16% (Table 3). Nevertheless, all seven QTLs had a negative additive effect on the trait, meaning that increased height is associated with the Tift 454 allele. In the case of stem diameter, we identified four QTLs with two each on LG3 and LG7. Interestingly the QTLs for plant height and stem diameter are co-localized at 15.46 cM on LG3 flanked by the two SNP markers S3_3095 and S3_3262. Like plant height, the QTL identified for stem diameter also had the negative additive effect on the trait.

QTLs for yield related traits- days to heading, panicle diameter, panicle length, grain yield, 1000 seed weight

Days to heading, panicle diameter, panicle length, grain yield per head and 1000 seed weight are the key yield related traits that directly impact grain yield in pearl millet. We identified six QTLs for days to heading and 11 QTLs for panicle diameter and seven QTLs for panicle length, four QTLs for grain yield and two QTLs for 1000 seed weight (Table 3). Further, among these QTLs several identified were overlapping and consistent across different years. For example, a QTL for days to heading on LG5 at 0.48 cM was identified in 2006 and 2013, demonstrating the consistency of this trait across years. No QTLs were detected for days to heading based on 2007 phenotypic evaluations of the RIL population. The QTLs for days to heading explained 8.38 to 14.22% of the phenotypic variation for the trait. In case of panicle diameter, we identified seven QTLs in 2007 on LG1, 2 and 3, of which four detected on LG3 were consistent with QTLs detected in 2013. The phenotypic variation explained for panicle diameter ranged from 6.28 to 21.60%. In the case of panicle length, seven QTLs were detected on LG3 and LG7 for 2007 and 2013 evaluations. Panicle diameter and panicle length shared four common QTLs on LG3 over the years. Further among these eight QTLs detected for panicle length and panicle diameter, four QTLs were based on 2007 evaluations and four based on 2013 (Figure 1). However, no QTLs were identified based on 2006 evaluation data for both the traits. For panicle length, five QTLs were detected in 2007 on LG3 and LG7, of which two overlapped with the 2013 evaluations and the PVE ranged from 6.01 to 26.99%. There were five genomic positions on LG2 (9.64 cM) for days to heading and panicle diameter and LG3 (0.01 cM, 9.95 cM, 14.8 cM, 15.46 cM) for panicle diameter and panicle length that were consistently identified in 2007 and 2013 with the same marker interval having the highest LOD and PVE. Furthermore, most of the QTLs for both panicle length and diameter had negative additive effects on the traits, meaning that greater panicle length and diameter were associated with Tift 454 alleles. In the case of grain yield, four QTLs were detected, of which two (at 27.52 cM) overlapped for 2006 and 2007. Two QTLs were detected for 1000 seed weight in 2013, of which one overlapped with panicle length and stem diameter on LG7 at 21.03 cM (Table 3).

QTLs for *Pyricularia* leaf spot disease

We identified one QTL on LG5 at 50.7 cM between S5_7467 and S5 7482 with 9.23% PVE and a negative additive effect of 0.58. Since a higher number indicates greater disease, resistance is associated with the Tift 99B allele.

QTLs for nematode resistance

We evaluated the RIL population for root-knot nematode resistance by rating egg masses in greenhouse experiments in 2009. All three replicates were analyzed, as well as replicate 1 and 2, or just replicate 3 due to difference in phenotyping of replicate 3 as compared to 1 and 2. Three QTLs on LG2 explained 23.84-32.51% of phenotypic variation were identified (Table 3, Figure 1). Notably, QTLs identified in the present study are the major QTLs located between S2_8057 and S2_8333, and have positive additive effects on the trait, meaning that resistance is associated with the Tift 454 allele.

10.3389/fpls.2025.1588485

TABLE 3 Summary of QTLs identified for morphological and yield-related traits based on phenotype of the pearl millet RIL population (Tift99B × Tift454) during 2006, 2007, 2009 and 2013.

Trait	Year of phenotyping	Linkage group	Position (cM)	Left marker	Right marker	LOD	PVE (%)	Additive
Plant Height (cm)	2006	2	9.64	S2_1572	S2_1945	3.07	6.38	-0.03
	2006	2	39.12	S2_6939	S2_6955	3.06	6.70	-0.03
	2013	5	0.01	S5_0001	S5_0012	3.54	7.20	-0.02
	2013	5	4.97	S5_0720	S5_0743	5.11	10.16	-0.02
	2013	3	14.8	S3_2814	S3_3095	4.69	9.58	-0.02
	2013	3	15.46	S3_3095	S3_3262	4.89	10.02	-0.03
	2013	3	26.14	S3_6344	S3_6554	3.91	8.27	-0.02
Stem Diameter (mm)	2013	3	15.46	S3_3095	S3_3262	4.68	9.83	-0.05
	2013	3	16.46	S3_3095	S3_3262	5.04	10.59	-0.05
	2013	7	24.1	S7_2292	S7_2386	5.99	13.25	0.06
	2013	7	21.03	S7_1713	S7_1714	7.74	16.99	0.07
Days To Heading (d)	2006	5	0.01	S5_0001	S5_0012	6.43	13.43	0.02
	2006	5	0.48	S5_0012	S5_0062	6.86	14.22	0.02
	2013	5	0.01	S5_0001	S5_0012	4.93	10.68	0.01
	2013	5	0.48	S5_0012	S5_0062	4.76	10.34	0.01
	2013	2	9.64	S2_1572	S2_1945	3.74	8.38	-0.01
	2013	2	14.17	S2_2218	S2_2320	5.34	11.98	-0.01
Panicle Diameter (mm)	2007	1	36.19	S1_4037	S1_4225	4.55	8.09	0.02
	2007	2	9.64	S2_1572	S2_1945	3.76	6.29	-0.01
	2007	2	11.42	S2_1952	S2_2171	3.92	6.54	-0.01
	2007	3	0.01	S3_0001	S3_0146	4.60	9.69	-0.02
	2007	3	9.95	S3_2072	S3_2266	8.87	17.99	-0.02
	2007	3	14.8	S3_2814	S3_3095	11.10	21.60	-0.03
	2007	3	15.46	S3_3095	S3_3262	10.44	20.71	-0.03
	2013	3	0.01	S3_0001	S3_0146	3.24	6.29	-0.03
	2013	3	9.95	S3_2072	S3_2266	5.88	10.56	-0.05
	2013	3	14.8	S3_2814	S3_3095	6.12	10.79	-0.05
	2013	3	15.46	S3_3095	S3_3262	5.84	10.41	-0.05
Panicle Length (cm)	2007	7	21.03	S7_1713	S7_1714	3.03	6.02	0.02
	2007	3	0.01	S3_0001	S3_0146	3.56	6.77	-0.02
	2007	3	9.95	S3_2072	S3_2266	11.58	27.00	-0.04
	2007	3	14.8	S3_2814	S3_3095	8.97	22.78	-0.03
	2007	3	15.46	S3_3095	S3_3262	8.02	20.43	-0.03
	2013	3	14.8	S3_2814	S3_3095	4.97	10.20	-0.03
	2013	3	15.46	S3_3095	S3_3262	4.89	10.14	-0.03

(Continued)

10.3389/fpls.2025.1588485

TABLE 3 Continued

Trait	Year of phenotyping	Linkage group	Position (cM)	Left marker	Right marker	LOD	PVE (%)	Additive
Grain Yield (g/head)	2006	6	26.77	S6_4678	S6_4837	5.95	16.02	0.13
	2006	6	27.52	S6_4837	S6_4897	5.07	13.47	0.12
	2007	6	27.52	S6_4837	S6_4897	3.45	7.08	0.10
	2007	5	7.47	S5_1420	S5_1614	4.89	10.22	-0.10
1000 Seed Weight (g)	2013	7	21.03	S7_1713	S7_1714	4.09	10.52	0.02
	2013	7	24.1	S7_2292	S7_2386	4.56	11.50	0.02
Pyricularia Leaf Spot	2013	5	50.7	S5_7467	S5_7482	3.32	9.23	-0.58
Egg mass Avg trial (1,2&3)	2009	2	50.99	S2_8057	S2_8319	10.72	28.37	1.12
Egg mass trial (1&2)	2009	2	51.06	S2_8319	S2_8326	11.78	32.51	0.13
Egg mass trial 3	2009	2	51.42	S2_8326	S2_8333	8.94	23.84	0.14

Candidate genes for different traits

A total of 19,762 genes were identified between the markers flanking 45 QTL regions identified in this study. Of these, 13,214 genes are identified in the major QTL regions (Supplementary Table S5). And also among these, 6884 are unique genes in these QTL regions (Supplementary Tables S6, S7). A large number of unique genes were identified in the grain yield QTL region (4091 genes). A set of 294, 408, 868, 1045, 1124 and 1526 unique genes were identified for stem diameter, RKN egg mass, panicle length, days to heading, panicle diameter and plant height respectively. We identified only one gene for *Pyricularia* leaf spot. Interestingly all

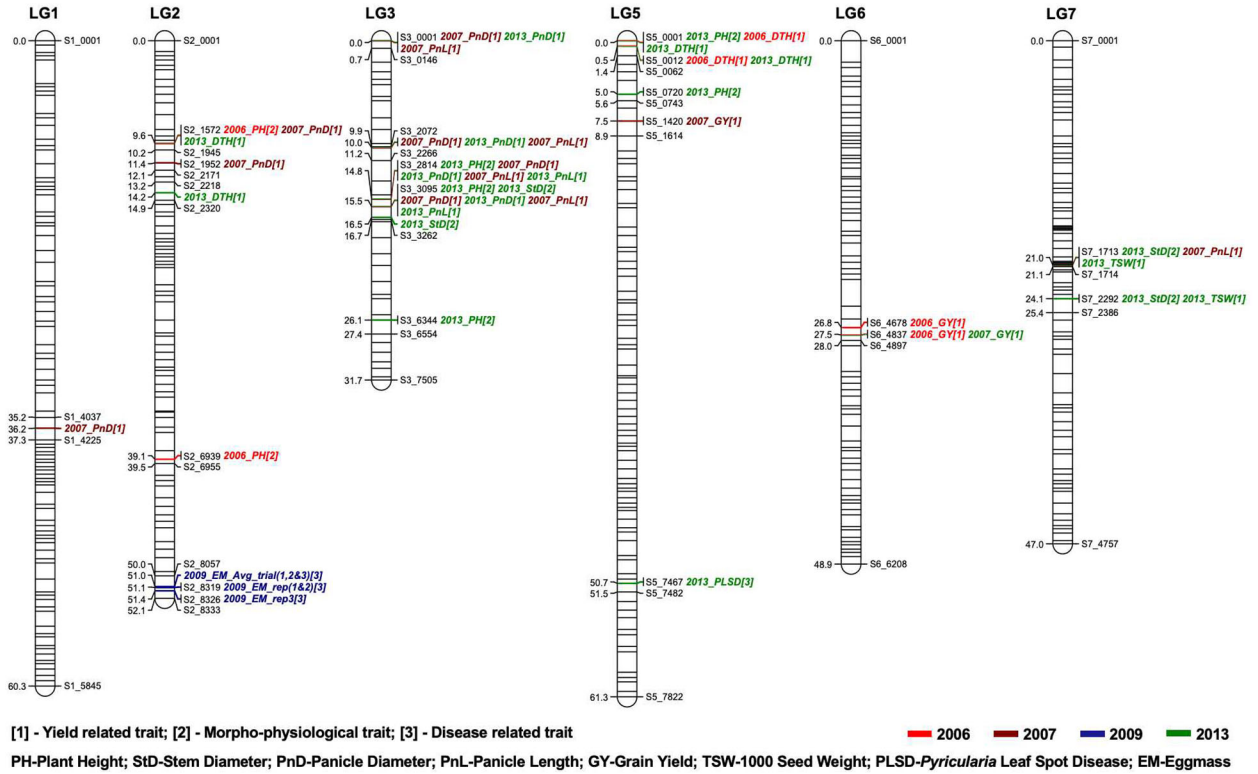


FIGURE 1
Genetic map constructed using RIL populations derived from Tift $99D_2B_1$ × Tift 454 comprise 505 SNP markers. Number of markers mapped ranges from 35 to 89 per linkage group. Linkage distances are described with a scale in cM on the left side and the QTLs regions are given on the right side of each bar.

127 and 1526 genes identified in the 1000 seed weight and plant height QTL regions are unique. The known relevant genes for these traits were extracted with Gene Ontology (GO) terms. The genes Pgl_GLEAN_10004502 and Pgl_GLEAN_10018466, that encode a heat shock protein Hsp90 are present in the QTL regions of days to heading (Supplementary Tables S6, S7). Among these 19,762 genes identified in the QTL regions, we summarized the number of genes in each QTL region for a given trait. Further, on functional annotation, 9057 genes had no reported function, and the remaining 10,705 genes encode for known functional proteins. We have focused discussion only on those genes that are reported to have influenced the trait in model species or known crop plants or that could be deployed in breeding programs for pearl millet improvement. These genes are also prioritized based on existing literature and previously characterized genes.

Discussion

Pearl millet is a resilient, drought-tolerant crop that flourishes in harsh environments. Historically labeled as an orphan crop due to its limited genomic resources, recent advancements in next-generation sequencing (NGS) technologies have revolutionized genomic research in pearl millet. Previously, we developed a comprehensive set of 16,650 markers, including 1,191 core markers, which facilitated the detection of numerous quantitative trait loci (QTLs) that might have been missed in less saturated maps (Punnuri et al., 2016). In this study, we constructed a genetic map using 505 SNP markers derived from the core markers, excluding those with segregation distortion, to identify QTLs associated with morphological, disease resistance and yield-related traits in a RIL populations at two distinct locations.

The parental inbreds used for developing RILs in this study are Tift 454 and Tift 99B. Tift 454 and Tift 99A are the parents of TifGrain 102, with Tift 99B serving as a male fertile maintainer line for Tift 99A (Hanna et al., 2005a, 2005b). The parental line Tift 454 exhibits high resistance to the root-knot nematode species *Meloidogyne incognita* (Kofoid & White) Chitwood (Hanna et al., 2005a); Tift 99B is resistant to rust (*Puccinia substriata* var. *indica*) and other diseases such as leaf spot caused by *Pyricularia grisea*. Although the differences between the two parents are minimal for several traits, they fall within the expected ranges for cultivated grain-type pearl millet inbreds in the United States.

Our study revealed significant variation among the RILs for all traits examined. Traits with the highest heritability values (0.78-0.93) included 1000 seed weight (2013), grain yield (2006, 2007), panicle diameter (2007, 2013), stem diameter (2013), and *Pyricularia* leaf spot disease (2013). Similarly, high broad-sense heritability values ranging from 0.52 (panicle length) to 0.86 (plant height) were reported for panicle diameter and plant height (Kumar et al., 2021). Selecting traits with higher heritability is a powerful approach in genetics and breeding, as it leverages the genetic component of trait variation for more effective and predictable improvements.

In this study, we utilized GBS-based SNP markers to construct a high-density genetic map and identify QTL. Traditionally, QTL analyses in pearl millet have predominantly used SSR markers or DArT markers (Ambawat et al., 2016; Kumar et al., 2017, 2021). However, a few studies have employed GBS-based SNPs (Punnuri et al., 2016; Pucher et al., 2018) or SSCP-SNP markers (Bertin et al., 2005). Most of SSR-based genetic maps contained very few markers, for instance, a consensus map combining four different populations had only 174 loci (Rajaram et al., 2013). GBS-based maps often have several hundred markers (314 for Moumouni et al., 2015; 460 for Pucher et al., 2018; and 505 for the current study). High-density genetic maps enable identification of precise genomic regions for use in breeding programs for trait improvement (Punnuri et al., 2024).

All the traits were not measured across all years and locations, but only four traits are common across two seasons and locations. Although a prior study showed that eight QTLs explained 42.7% of observed phenotypic variation for panicle length and five QTLs explained 45.8% of observed phenotypic variation (Vengadessan et al., 2013), our research identified 29 major QTLs (explaining >10% of PVE each) for traits such as plant height, stem diameter, days to heading, panicle diameter, panicle length, grain yield, 1000 seed weight, leaf spot resistance, and root-knot nematode egg masses. Notably, the total PVE for egg mass trial (1&2), egg mass average trial (1,2&3) and egg mass trial 3 were 32.51%, 28.37%, and 23.84%, respectively. Additionally, we identified 16 minor QTLs (explaining <10% PVE each) across all traits.

These QTLs were detected with high LOD values (ranging from 3.02 to 11.78) on all linkage groups (LGs) except LG4, explaining significant phenotypic variation for yield-related traits, including biotic resistance. Most of these QTLs were consistently expressed across different locations, years, and traits. LG4 showed less polymorphic markers associated with QTLs due to shared ancestry between the parents (Punnuri et al., 2016).

The highest number of QTLs (19) were detected on LG3, showing overlap among traits and years. For instance, plant height (2013), panicle diameter (2007 & 2013), and panicle length (2007 & 2013) were consistently observed at 14.8 cM on LG3 across different experimental locations. Similarly, plant height (2013), stem diameter (2013), panicle diameter (2007 & 2013), and panicle length (2007 & 2013) were identified on LG3 at 15.46 cM between markers S3_3095 and S3_3262. Panicle diameter (2007 & 2013) and panicle length (2007) were consistently found on LG3 at 0.01 cM and 9.95 cM between evaluations in Tifton and Fort Valley. Among these four traits (plant height, stem diameter, panicle length and panicle diameter), QTLs for two traits (panicle length and panicle diameter) had four overlapping QTL that were consistently identified at 0.01 cM, 9.95 cM, 14.8 cM, and 15.46 cM on LG3. These co-localized QTL could represent pleiotropic effects or tightly linked genes. Determining which of these is actually the case will require the development of fine-mapping populations to further break up the linkage blocks in this region, and may benefit from multi-trait QTL analyses.

Other notable QTLs included plant height (2006), days to heading (2013), and panicle diameter (2013) at 9.64 cM on LG2. LG2 also harbored the nematode resistance QTL between markers S2_8057 and S2_8333. Leaf spot caused by *Pyricularia* due to frequent rains and high humidity is a major concern in the

southeastern United States. The QTLs reported for leaf spot are different from our previous report (Punnuri et al., 2016) as we removed distorted markers and used composite interval mapping in the current study. Nevertheless, we report identification of QTL for *Pyricularia* leaf spot in this study confirming the presence of a genomic region governing *Pyricularia* leaf spot on LG5. Earlier one RAPD marker (OP-D11700, 5.6 cM) was reported to be linked to *Pyricularia* leaf spot resistance using a population derived from Tift85DB and Tift65 (Morgan et al., 1998).

Grain yield-related QTLs (2006 & 2007) were found on LG6 in close proximity with overlapping regions at 27.52 cM. Similar results were found for grain yield on LG6 in a study by Kumar et al. (2017). Panicle length (2007) was linked with 1000 seed weight and stem diameter on LG7. These results indicate that most yield and domestication-related traits, such as panicle morphology and disease resistance, were co-localized in clusters on specific genomic regions. These regions were likely selected as linkage blocks during domestication for yield improvement. Consequently, many traits remained undetected on LG4. This is likely due to both biological and technical reasons. Biologically, the parental lines Tift 454 and Tift 99B share a common ancestry and both possess the *dw2* dwarfing gene derived from Tift $23D_2$ (Punnuri et al., 2016), which resulted in reduced polymorphism on LG4. Technically, LG4 had fewer markers compared to most other linkage groups, which limited our ability to detect QTLs due to reduced resolution and statistical power.

Previous studies using multi-location phenotyping data on a RIL population derived from ICMB 841-P3 × 863B-P2 identified 13 QTLs for plant height, 11 QTLs for panicle length, and 14 QTLs for 1,000-seed weight (Kumar et al., 2017). Vengadessan et al. (2013) reported QTLs for panicle length and diameter on LG3. Additionally, for panicle length, five major QTLs were mapped on LG3 in the current study, which is consistent with the findings of Vengadessan et al. (2013). A minor QTL for panicle length was also identified on LG7 in this study, supporting previous reports by Vengadessan et al. (2013) and Kannan et al. (2014). Recently, Gupta et al. (2024) identified 34 Meta-QTL regions with a 3.63-fold reduction in confidence intervals, harboring genes for agronomic and yield-related traits (plant height, flowering time) and yield-related traits (panicle length, 1000 grain weight, seed yield per plant). In a genomic scanning and genome-wide association studies, 314 SNPs detected on chromosomes 1, 2 and 5 were found to be associated with flowering and agromorphologic traits, of which the majority of them were linked with panicle length and width (Faye et al., 2022). Also, it was shown that presence of the alternate allele on SNP Chr3_at 18,486,380 bp, increased 1000 seed weight. In other studies, major and stable QTL associated with flowering, number of productive tillers, ear head length, and test weight were mapped on chromosomes 1 and 3 using a RIL population derived from a cross between PT6029 and PT6129 (Subbulakshmi et al., 2024). Similar to Kannan et al. (2014), in the current study a QTL for days to heading was mapped on LG5 for 2006 and 2013 phenotyping years.

In this study we reported consistent and stable QTLs for panicle diameter and panicle length that shared the same genomic regions on LG3; however, the meta-QTL analysis performed by Gupta et al. (2024) based on 12 independent studies for various traits harbored QTLs and genes for plant height, 1000 seed weight, panicle length, seed yield per plant, and flowering time in the Meta-QTL1.1 on LG1. Nevertheless, QTLs for panicle length and panicle diameter were present in Meta-QTL3.6 on LG3. The co-localized QTLs reported in our study can be used in marker assisted introgression for simultaneous improvement of several traits. Grain number and grain mass are critical yield components that directly influence grain yield (Kumar et al., 2017). Panicle length significantly affects panicle weight, which is a key factor in calculating the panicle harvest index. QTLs for mean grain yield across various moisture environments have been mapped to linkage groups LG2, LG3, and LG4 under different post-flowering moisture conditions (Bidinger et al., 2007). In the current study, QTLs for plant height, panicle length, panicle diameter, and stem diameter were localized on LG3, supporting previous findings. A QTL for 1000 seed weight on LG7 was previously reported in two different mapping populations (Kumar et al., 2021, 2017), similar to our findings but at different positions on LG7.

A recent study using a RIL population derived from ICMS 8511-S1-17-2-1-1-B-P03 × AIMP 92901-S1-183-2-2-B-08 identified 15 QTLs for plant height scattered across seven chromosomes, five QTLs for panicle traits dispersed across four chromosomes, and two QTLs for thousand grain weight spanning LG3 and LG7 (Kumar et al., 2021). The co-localization of QTLs offers significant advantages in pearl millet breeding by enabling the simultaneous improvement of multiple traits or leveraging the selection of one trait to enhance others. In this study, we also identified stable QTLs for traits such as days to heading, panicle diameter, and panicle length that were consistently mapped to the same positions across different years. These stable QTLs can be utilized in marker-assisted selection to improve these traits cost-effectively.

Several studies have identified key yield-related genes in pearl millet, shedding light on their chromosomal positions and functional roles. Four genes, (Pgl_GLEAN_10031557, Pgl_GLEAN_10008690, Pgl_GLEAN_10007548 and Pgl_GLEAN_10005530) identified in this study encoding phospholipase were identified in the panicle diameter and grain yield QTL regions are believed to be responsible for determining pearl millet yield. The results from earlier studies further confirm the role of phospholipase gene on grain yield of pearl millet. For instance, a genome-wide analysis of the phospholipase gene family revealed 44 genes distributed across the seven chromosomes, with the highest concentration on chromosome 1 (12 genes), followed by chromosome 5 (8 genes), chromosome 6 (7 genes), and fewer on the remaining chromosomes. These phospholipase genes are implicated in various physiological processes, influencing stress responses that ultimately affect yield (Moin et al., 2024). Flowering time, a crucial determinant of yield in pearl millet, is regulated by several genes located primarily on chromosome 2. Among them, *PhyC* (PHYTOCHROME C) plays a pivotal role in the photoperiodic induction of flowering, with polymorphisms near this gene linked to variations in flowering time and yield components. Similarly, *FRS12* (FAR1-RELATED SEQUENCE 12) and *Hd3a* (HEADING DATE 3a) have been identified as key regulators of the transition from vegetative to reproductive stages. We identified a gene Pgl_GLEAN_10007711 in

the panicle diameter QTL that encodes for FAR1. Additionally, *OsHAC1* (HISTONE ACETYLTRANSFERASE 1), also found on chromosome 2, is involved in chromatin remodeling, affecting gene expression related to flowering (Faye et al., 2022). We identified the gene Pgl_GLEAN_10010455 in grain yield QTL region encodes for histone acetyltransferase. Beyond flowering regulation, drought tolerance is another critical factor influencing yield stability in pearl millet. A comprehensive study identified 74 drought-associated genes distributed across all seven chromosomes, with chromosome 1 containing the highest number (14 genes), followed by chromosomes 6 (12 genes), 4 (11 genes), and others with fewer genes. These drought-responsive genes function in various pathways that enhance resilience under water-limited conditions, contributing to yield improvement under abiotic stress (Chakraborty et al., 2022).

In addition, genes that are indirectly related to yield were also reported, for instance, *PmAAP1* (Amino Acid Permease 1) and *PmGS1* (Glutamine Synthetase 1) play crucial roles in nitrogen assimilation and amino acid transport, which indirectly contribute to yield by improving grain nutritional quality (Singh et al., 2024b). In our study, the gene Pgl_GLEAN_10003392 present on chromosome 7 encodes for glutamine synthetase and is present in the grain yield QTL region. Similarly, the gene Pgl_GLEAN_10022286 present on chromosome 7 in the grain yield QTL region encodes for amino acid permease. Furthermore, resistance to foliar blast, a major disease affecting pearl millet yield, has been linked to several genes identified in a GWAS of 281 inbred lines (Singh et al., 2024a). Among them, *PmNPR1* (Nonexpressor of Pathogenesis-Related Genes 1) and *PmWRKY45* (WRKY Transcription Factor 45) were found to be key regulators of blast resistance, mainly located on chromosomes 1, 2, and 6. These genes enhance disease resistance by modulating defense pathways, ultimately stabilizing yield under biotic stress (Singh et al., 2024b). In contrast, in this study, we identified Pgl_GLEAN_10014680 on chromosome 7 within a grain yield QTL region, which is annotated as a pathogenesis-related gene. While pathogenicity is a trait typically associated with pathogens, host plants also harbour genes that influence disease outcomes, either by enhancing susceptibility or activating defence responses. The presence of Pgl_GLEAN_10014680 in a yield-associated QTL suggests a potential link between disease response and productivity, warranting further functional validation to determine its role in host-pathogen interactions. Additionally, rhizosheath formation, which influences water and nutrient uptake in pearl millet, has been linked to *PmEXPA1* (Expansin A1) and *PmMATE37* (Multidrug And Toxic Compound Extrusion 37). In this study, we identified the gene Pgl_GLEAN_10006620 in the grain yield QTL region that encodes for expansin. These genes regulate root-soil interactions and help plants maintain productivity in arid conditions, indirectly enhancing yield stability in drought-prone environments (de la Fuente Cantó et al., 2022). In the search of candidate genes for *Pyricularia* QTL on LG5, we identified a gene Pgl_GLEAN_10037626 that encodes for tetratricopeptide repeat (TPR) domains that are structural motifs comprising tandem repeats of 34 amino acids that facilitate protein-protein interactions (Blatch and Lässle, 1999). In cereals, TPR-containing proteins play significant roles in various biological processes, including disease resistance (Schapire et al., 2006). For instance, in rice, the TPR-domain RNA-binding protein BSR-K1 has been identified as a negative regulator of broad-spectrum disease resistance. A loss-of-function mutation in the *Bsr-k1* gene results in enhanced resistance to multiple pathogens, including *Magnaporthe oryzae* (anamorph: *Pyricularia oryzae*) and *Xanthomonas oryzae* pv. *oryzae* (Zhou et al., 2018). This mutation leads to the production of a truncated BSR-K1 protein that cannot bind to the mRNAs of phenylalanine ammonia-lyase (PAL) genes, resulting in increased accumulation of PAL transcripts. The elevated PAL levels enhance the biosynthesis of secondary metabolites like lignin, thereby strengthening the plant's defense mechanisms without adversely affecting yield (Zhou et al., 2018). Additionally, another TPR-domain protein in rice, OsTPR1, has been shown to bind with a chitinase from *Magnaporthe oryzae* thus allowing the plant to exhibit an immune response (Yang et al., 2019). Similarly in wheat, TPR-containing proteins have been implicated in regulating agronomic traits that may influence disease resistance. For instance, the TPR protein TaTPR-B1 has been shown to regulate spike compactness, a trait that can affect susceptibility to certain pathogens (Zhu et al., 2025). While the direct role of TaTPR-B1 in disease resistance requires further investigation, its influence on spike architecture suggests a potential indirect effect on pathogen interactions. Hence, the gene may be responsible for *Pyricularia* leaf spot resistance in pearl millet.

Meloidogyne incognita are plant-parasitic nematodes that cause significant damage to pearl millet in the southern states where it is grown. Additionally, their presence in the soil affects rotational crops, especially peanuts (*Arachis hypogaea* L.) and cotton (*Gossypium hirsutum* L.). We identified major QTLs for resistance to root-knot nematode (RKN) egg masses on LG2 between markers S2_8057 and S2_8333. Previously, a major QTL for resistance to southern root-knot nematode, QMi-LG2, with a LOD score of 14, explained 32.0% of the phenotypic variance was also mapped to LG2 using an AFLP and SSR-based genetic map (Liu, 2012).

Interestingly, in this study using the same phenotyping data, we report three major QTLs on LG2 with high LOD scores (>8), explaining 32.51%, 28.37%, and 23.84% of the phenotypic variance, indicating a strong association of root-knot nematode egg masses on LG2. The heritability values for nematode and leaf spot resistance were also high. For nematode resistance, the RKN resistance gene from Tift 454 is incompletely dominant or semidominant. The parent Tift 454 is almost immune, the F_1 generation is resistant but still has egg masses, and the parent Tift 99B is susceptible. The ancestors of Tift 454 include Tift $23D_2A_1$ and napiergrass (*Pennisetum purpureum* Schumacher) (Hanna et al., 2005a). Previous studies have suggested that the nematode resistance gene may have originated from napiergrass (Liu, 2012). Among 408 unique genes identified in the egg mass QTL region, the gene Pgl_GLEAN_10018312, encodes for Thaumatin-like proteins (TLPs), the members of the pathogenesis-related (PR) protein family, specifically classified under PR-5. They are recognized for their role in plant defense mechanisms against various pathogens, including fungi and bacteria (Sharma et al., 2022). Nevertheless, the gene Pgl_GLEAN_10018133 encodes for plant disease resistance response protein may be implicated for nematode resistance.

Initial scanning for total number of genes within QTL marker intervals yielded 19,762 genes within the QTL regions. Further

refining and annotation gave us around 10,705 annotated genes within these 45 QTL regions, including those coding for heat shock protein Hsp90 (Supplementary Tables S6, S7). Previously, 28 Hsp20 genes were identified based on transcriptome profiling of pearl millet under high-temperature stress (Mukesh Sankar et al., 2021). Hsp90 plays crucial roles in protein folding, stress response, and developmental processes, which are vital for coping with heat and drought stress, making it a target for improving crop resilience.

In summary, next-generation sequencing (NGS) technologies play a prominent role in improving linkage map saturation and its utility in breeding programs. The major QTLs identified in this study and the genes reported in these QTL regions can be used in genomics-assisted breeding to enhance key agronomic and yield-related traits in pearl millet.

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

SV: Data curation, Formal analysis, Resources, Visualization, Writing – review & editing, Software. JK: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Writing – original draft, Writing – review & editing. AS: Data curation, Resources, Writing – review & editing. RN: Data curation, Resources, Writing – review & editing. LL: Data curation, Formal analysis, Investigation, Methodology, Resources, Writing – review & editing. PC: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Supervision, Writing – review & editing, Project administration. RK: Resources, Writing – review & editing. BF: Formal analysis, Resources, Software, Writing – review & editing. MT: Data curation, Formal analysis, Resources, Writing – original draft, Writing – review & editing. PT: Formal analysis, Methodology, Resources, Supervision, Validation, Writing – review & editing. KH: Data curation, Formal analysis, Software, Validation, Writing – original draft, Writing – review & editing, Investigation, Resources. JW: Data curation, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. HS: Data curation, Writing – review & editing, Resources. BS: Conceptualization, Funding acquisition, Project administration, Supervision, Writing – original draft, Writing – review & editing, Investigation. SP: Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, 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. We acknowledge USDA NIFA Evans-Allen grant number GEOX 5221-336129 and multi-state project 336154. We acknowledge previous funding support received by BS for capacity building grant USDA-National Institute of Food and Agriculture-Grant#GEOX-2008-02595.

Conflict of interest

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

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

Generative AI statement

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

Publisher's note

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.2025.1588485/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Linkage map along with identified QTL regions.

References

Agarwal, P., Singh, B. R., Gajbe, U., Kalambe, M. A., and Bankar, M. (2023). Managing diabetes mellitus with millets: A new solution. *Cureus* 15, e44908. doi: 10.7759/cureus.44908

Ambawat, S., Senthilvel, S., Hash, C. T., Nepolean, T., Rajaram, V., Eshwar, K., et al. (2016). QTL mapping of pearl millet rust resistance using an integrated DArT-and SSR-based linkage map. *Euphytica* 209, 461–476. doi: 10.1007/s10681-016-1671-9

Andrews, D. J., and Kumar, K. A. (1992). Pearl millet for food, feed, and forage. *Adv. Agronomy* 48, 89–139. doi: 10.1016/S0065-2113(08)60936-0

Azhaguvel, P. (2001). Linkage map construction and identification of QTLs for downy mildew (*Sclerospora graminicola*) resistance in pearl millet [*Pennisetum glaucum* (L.) R. Br.], Ph.D. Thesis. TNAU, Coimbatore.

Bernardo, R. (2013). Genome-wide markers for controlling background variation in association mapping. *Plant Genome* 6, 1–9. doi: 10.3835/plantgenome2012.11.0028

Bertin, I., Zhu, J. H., and Gale, M. D. (2005). SSCP-SNP in pearl millet—a new marker system for comparative genetics. *Theor. Appl. Genet.* 110, 467–1472. doi: 10.1007/s00122-005-1981-0

Bidinger, F. R. T., Nepolean, C. T., Hash, R. S., Yadav,, and Howarth, C. J. (2007). Quantitative trait loci for grain yield in pearl millet under variable post flowering moisture conditions. *Crop Sci.* 47, 969–980. doi: 10.2135/cropsci2006.07.0465

Blatch, G. L., and Lässle, M. (1999). The tetratricopeptide repeat: a structural motif mediating protein-protein interactions. *BioEssays* 21, 932–939. doi: 10.1002/(SICI)1521-1878(199911)21:11<932::AID-BIES5>3.0.CO;2-N

Burgarella, C., Cubry, P., Kane, N. A., Varshney, R. K., Mariac, C., Liu, X., et al. (2018). A Western Sahara centre of domestication inferred from pearl millet genomes. *Nat. Ecol. Evol.* 2, 1377–1380. doi: 10.1038/s41559-018-0643-y

Chakraborty, A., Viswanath, A., Malipatil, R., Semalaiyappan, J., Shah, P., Ronanki, S., et al. (2022). Identification of candidate genes regulating drought tolerance in pearl millet. *Int. J. Mol. Sci.* 23, 6907. doi: 10.3390/ijms23136907

Chelpuri, D., Sharma, R., Durga, K. K., Katiyar, P., Mahendrakar, M. D., Singh, R. B., et al. (2019). Mapping quantitative trait loci (QTLs) associated with resistance to major pathotype-isolates of pearl millet downy mildew pathogen. *Eur. J. Plant Pathol.* 154 (4), 983–994. doi: 10.1007/s10658-019-01718-xhemisquy

Davis, A. J., Dale, N. M., and Ferreira, F. J. (2003). Pearl millet as an alternative feed ingredient in broiler diets. *J. Appl. Poultry Res.* 12, 137–144. doi: 10.1093/japr/12.2.137

de la Fuente Cantó, C., Diouf, M. N., Ndour, P. M. S., Debieu, M., Grondin, A., Passot, S., et al. (2022). Genetic control of rhizosheath formation in pearl millet. *Sci. Rep.* 12 (1), 9205. doi: 10.1038/s41598-022-13234-w

FAOSTAT (2020). *FAOSTAT Online Database* (Rome, Italy: FAO).

Faye, A., Barnaud, A., Kane, N. A., Cubry, P., Mariac, C., Burgarella, C., et al. (2022). Genomic footprints of selection in early-and late flowering pearl millet landraces. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.88063

Gangashetty, P. I., Motagi, B. N., Pavan, R., and Roodagi, M. B. (2016). Breeding crop plants for improved human nutrition through biofortification: progress and prospects. *Adv. Plant Breed. Strategies: Agronom. Abiotic Biotic. Stress Traits* pp. 35-76.

Gupta, S., Rangari, S. K., and Sahu, A. (2024). Meta-QTL analysis reveals the important genomics regions for biotic stresses, nutritional quality and yield related traits in pearl millet. *CABI Agric. Biosci.* 5 (1), p. 36. doi: 10.1186/s43170-024-00230-5

Hancock, D., Kichler, J., Smith, B., and Hicks, R. (2018). Georgia forages: grass species. (Georgia: University of Georgia Cooperative Extension) 1351.

Hanna, W. W., Dove, R., Hill, G. M., and Smith, R. (1991). "Pearl millet as an animal feed in the US," in *American Society of Agronomy Abstracts, Southern Branch*(Fort Worth, Texas).

Hanna, W. W., and Wells, H. D. (1989). Inheritance of *Pyricularia* leaf spot resistance in pearl millet. *J. Hered.* 80, 145–147. doi: 10.1093/oxfordjournals.jhered.a110814

Hanna, W., Wilson, J., and Timper, P. (2005a). Registration of pearl millet parental line Tift 454. *Crop Sci.* 45, 2670–2672. doi: 10.2135/cropsci2005.0171

Hanna, W., Wilson, J., and Timper, P. (2005b). Registration of pearl millet parental lines Tift 99 [D. sub. 2][A. sub. 1]/[B. sub. 1. *Crop Sci.* 45, 2671–2672. doi: 10.2135/cropsci2005.0172

Holbrook, C. C., Knauft, D. A., and Dickson, D. W. (1983). A technique for screening peanut for resistance to Meloidogyne arenaria. *Plant Disease* 67, 957–958. doi: 10.1094/PD-67-957

Kannan, B., Senapathy, S., Bhasker Raj, A. G., Chandra, S., Muthiah, A., Dhanapal, A. P., et al. (2014). Association analysis of SSR markers with phenology, grain, and stover-yield related traits in pearl millet (*Pennisetum glaucum* (L.) R. Br.). *Sci. World J.* 2014, 562327. doi: 10.1155/2014/562327

Kumar, S., Hash, C. T., Nepolean, T., Satyavathi, C. T., Singh, G., Mahendrakar, M. D., et al. (2017). Mapping QTLs controlling flowering time and important agronomic traits in pearl millet. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.01731

Kumar, S., Hash, C. T., Singh, G., Nepolean, T., and Srivastava, R. (2021). Mapping QTLs for important agronomic traits in an Iniadi-derived immortal population of pearl millet. *Biotechnology Notes.* 2, 26–32. doi: 10.1016/j.biotno.2021.06.001

Liu, L. (2012). *Genetic mapping and quantitative trait locus (QTL) analysis of root-knot nematode resistance in pearl millet. Ph.D. dissertation* Vol. 3 (Athens, GA, USA: University of Georgia).

Moin, M., Bommineni, P. R., and Tyagi, W. (2024). Exploration of the pearl millet phospholipase gene family to identify potential candidates for grain quality traits. *BMC Genomics* 25, 581. doi: 10.1186/s12864-024-10504-x

Morgan, R. N., Wilson, J. P., and Hanna, W. W. (1998). Molecular markers for rust and *Pyricularia* leaf spot disease resistance in pearl millet. *Theor. Appl. Genet.* 96, 413–420. doi: 10.1007/s001220050757

Moumouni, K. H., Kountche, B. A., Jean, M., Hash, C. T., Vigouroux, Y., Haussmann, B. I. G., et al. (2015). Construction of a genetic map for pearl millet, *Pennisetum glaucum* (L.) R. Br., using a genotyping-by-sequencing (GBS) approach. *Mol. Breeding* 35, 5.

Mukesh Sankar, S., Tara Satyavathi, C., Barthakur, S., Singh, S. P., Bharadwaj, C., and Soumya, S. L. (2021). Differential modulation of heat-inducible genes across diverse genotypes and molecular cloning of a sHSP from pearl millet [*Pennisetum glaucum* (L.) R. Br. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.659893

Parvathaneni, R. K., Jakkula, V., Padi, F. K., Faure, S., Nagarajappa, N., Pontaroli, A. C., et al. (2013). Fine-mapping and identification of a candidate gene underlying the *d2* dwarfing phenotype in pearl millet, *Cenchrus americanus* (L.) Morrone. *G3: Genes Genomes Genet.* 3, 563–572. doi: 10.1534/g3.113.005587

Passot, S., Gnacko, F., Moukouanga, D., Lucas, M., Guyomarc'h, S., Ortega, B. M., et al. (2016). Characterization of pearl millet root architecture and anatomy reveals three types of lateral roots. *Front. Plant Sci.* 7, 829. doi: 10.3389/fpls.2016.00829

Pei, J., Umapathy, V. R., Vengadassalapathy, S., Hussain, S. F. J., Rajagopal, P., Jayaraman, S., et al. (2022). A review of the potential consequences of pearl millet (*Pennisetum glaucum*) for diabetes mellitus and other biomedical applications. *Nutrients* 14, 2932. doi: 10.3390/nu14142932

Poncet, V., Lamy, F., Devos, K. M., Gale, M. D., Sarr, A., and Robert, T. (2000). Genetic control of domestication traits in pearl millet (*Pennisetum glaucum* L., Poaceae). *Theor. Appl. Genet.* 100, 147–159. doi: 10.1007/s001220050020

Pucher, A., Hash, C. T., Wallace, J. G., Han, S., Leiser, W. L., and Haussmann, B. I. (2018). Mapping a male-fertility restoration locus for the A 4 cytoplasmic-genic male-sterility system in pearl millet using a genotyping-by-sequencing-based linkage map. *BMC Plant Biol.* 18, 1–11. doi: 10.1186/s12870-018-1267-8

Punnuri, S., Wallace, J., and Knoll, J. (2024). "High-density genotyping for pearl millet linkage map improvement with next-generation sequencing data," in *The Pearl Millet Genome* (Springer International Publishing, Cham), 97–105.

Punnuri, S. M., Wallace, J. G., Knoll, J. E., Hyma, K. E., Mitchell, S. E., Buckler, E. S., et al. (2016). Development of a high-density linkage map and tagging leaf spot resistance in pearl millet using genotyping-by-sequencing markers. *Plant Genome* 9 (2). doi: 10.3835/plantgenome2015.10.0106

Rai, K. N., Gowda, C. L. L., Reddy, B. V. S., and Sehgal, S. (2008). Adaptation and potential uses of sorghum and pearl millet in alternative and health foods. *Compr. Rev. Food Sci. Food Saf.* 7, 320–396. doi: 10.1111/j.1541-4337.2008.00049.x

Rajaram, V., Nepolean, T., Senthilvel, S., Varshney, R. K., Vadez, V., Srivastava, R. K., et al. (2013). Pearl millet [*Pennisetum glaucum* (L.) R. Br.] consensus linkage map constructed using four RIL mapping populations and newly developed EST-SSRs. *BMC Genomics* 14, 1–16. doi: 10.1186/1471-2164-14-159

Satyavathi, C. T., Ambawat, S., Khandelwal, V., and Srivastava, R. K. (2021). Pearl millet: A climate-resilient nutricereal for mitigating hidden hunger and provide nutritional security. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.659938

Schapire, A. L., Valpuesta, V., and Botella, M. A. (2006). TPR proteins in plant hormone signaling. *Plant Signal Behav.* 1, 229–230. doi: 10.4161/psb.1.5.3491

Sehgal, D., Rajaram, V., Armstead, I. P., Vadez, V., Yadav, Y. P., Hash, C. T., et al. (2012). Integration of gene-based markers in a pearl millet genetic map for identification of candidate genes underlying drought tolerance quantitative trait loci. *BMC Plant Biol.* 12, 9. doi: 10.1186/1471-2229-12-9

Sharma, P. C., Sehgal, D., Singh, D., Singh, G., and Yadav, R. S. (2011). A major terminal drought tolerance QTL of pearl millet is also associated with reduced salt uptake and enhanced growth under salt stress. *Mol. Breeding* 27, 207–222. doi: 10.1007/s11032-010-9423-3

Sharma, A., Sharma, H., Rajput, R., Pandey, A., and Upadhyay, S. K. (2022). Molecular characterization revealed the role of thaumatin-like proteins of bread wheat in stress response. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.807448

Singh, S., Prakash, G., Nanjundappa, S., Malipatil, R., Kalita, P., Satyavathi, T. C., et al. (2024a). Novel SNPs linked to blast resistance genes identified in pearl millet through genome-wide association models. *Int. J. Mol. Sci.* 25, 12048. doi: 10.3390/ijms252212048

Singh, S., Yadav, C. B., Lubanga, N., Hegarty, M., and Yadav, R. S. (2024b). Genome-wide SNPs and candidate genes underlying the genetic variations for protein and amino acids in pearl millet (*Pennisetum glaucum*) germplasm. *Planta* 260, 63. doi: 10.1007/s00425-024-04495-y

Srivastava, R. K., Bollam, S., Pujarula, V., Pusuluri, M., Singh, R. B., Potupureddi., G., et al. (2020). Exploitation of heterosis in pearl millet: a review. *Plants* 9, 807. doi: 10.3390/plants9070807

Subbulakshmi, K., Karthikeyan, A., Murukarthick, J., Dhasarathan, M., Naveen, R., Sathya, M., et al. (2024). Consensus genetic linkage map and QTL mapping allow to capture the genomic regions associated with agronomic traits in pearl millet. *Planta* 260, 57. doi: 10.1007/s00425-024-04487-y

Thakur, R. P., Sharma, R., and Rao, V. P. (2011). Screening Techniques for Pearl Millet Diseases. *Technical Report* Vol. 1. Andhra Pradesh, India, International Crops Research Institute for the Semi-Arid Tropics.

Thudi, M., Palakurthi, R., Schnable, J. C., Chitikineni, A., Dreisigacker, S., Mace, E., et al. (2021). Genomic resources in plant breeding for sustainable agriculture. *J. Plant Physiol.* 257, 153351. doi: 10.1016/j.jplph.2020.153351

Timper, P., and Wilson, J. P. (2006). Root-knot nematode resistance in pearl millet from West and East Africa. *Plant Dis.* 90, 339–344. doi: 10.1094/PD-90-0339

Timper, P., Wilson, J. P., Johnson, A. W., and Hanna, W. W. (2002). Evaluation of pearl millet grain hybrids for resistance to *Meloidogyne* spp. and leaf blight caused by *Pyricularia grisea*. *Plant Dis.* 86, 909–914. doi: 10.1094/PDIS.2002.86.8.909

Tomar, M., Bhardwaj, R., Kumar, M., Singh, S. P., Krishnan, V., Kansal, R., et al. (2021). Nutritional composition patterns and application of multivariate analysis to evaluate indigenous Pearl millet (*Pennisetum glaucum* (L.) R. Br.) germplasm. *J. Food Composition Anal.* 103, 104086. doi: 10.1016/j.jfca.2021.104086

Varshney, R. K., Shi, C., Thudi, M., Mariac, C., Wallace, J., Zhang, H., et al. (2017). Pearl millet genome sequence provides a resource to improve agronomic traits in arid environments. *Nat. Biotechnol.* 35, 1–13. doi: 10.1038/nbt.3943

Varshney, R. K., Thudi, M., Pandey, M. K., Tardieu, F., Ojiewo, C., Vadez, V., et al. (2018). Accelerating genetic gains in legumes for the development of prosperous smallholder agriculture: integrating genomics, phenotyping, systems modelling and agronomy. *J. Exp. Bot.* 69, 3293–3312. doi: 10.1093/jxb/ery088

Vengadessan, V., Rai, K. N., Kannan Bapu, J. R., Hash, C. T., Bhattacharjee, R., Senthilvel, S., et al. (2013). Construction of genetic linkage map and QTL analysis of sink-size traits in pearl millet (*Pennisetum glaucum*). *Int. Scholarly Res. Notices* 2013 (1), p.471632. doi: 10.5402/2013/471632

Voorrips, R. E. (2002). MapChart: Software for the graphical presentation of linkage maps and QTLs. *J. Hered.* 93, 77–78. doi: 10.1093/jhered/93.1.77

Wang, S., Basten, C. J., and Zeng, Z. B. (2012). *Windows QTL Cartographer 2.5* (Raleigh, North Carolina, USA: Department of Statistics, North Carolina State University).

Wilson, J. P., and Hanna, W. W. (1992). Effects of gene and cytoplasm substitutions in pearl millet on leaf blight epidemics and infection by *Pyricularia grisea*. *Phytopathology* 82, 839–842. doi: 10.1094/Phyto-82-839

Yadav, O. P., Gupta, S. K., Govindaraj, M., Sharma, R., Varshney, R. K., Srivastava, R. K., et al. (2021). Genetic gains in pearl millet in India: Insights into historic breeding strategies and future perspective. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.645038

Yadav, R. S., Hash, C. T., Bidinger, F. R., Cavan, G. P., and Howarth, C. J. (2002). Quantitative trait loci associated with traits determining grain and stover yield in pearl millet under terminal drought-stress conditions. *Theor. Appl. Genet.* 104, 67–83. doi: 10.1007/s001220200008

Yadav, R. S., Hash, C. T., Bidinger, F. R., Devos, K. M., and Howarth, C. J. (2004). Genomic regions associated with grain yield and aspects of post flowering drought tolerance in pearl millet across stress environments and tester background. *Euphytica* 136, 265–277. doi: 10.1023/b:euph.0000032711.34599.3a

Yang, C., Yu, Y., Huang, J., Meng, F., Pang, J., Zhao, Q., et al. (2019). Binding of the *Magnaporthe oryzae* chitinase MoChia1 by a rice tetratricopeptide repeat protein allows free chitin to trigger immune responses. *Plant Cell.* 31, 172–188. doi: 10.1105/tpc.18.00382

Zhou, X., Liao, H., Chern, M., Yin, J., Chen, Y., Wang, J., et al. (2018). Loss of function of a rice TPR-domain RNA-binding protein confers broad-spectrum disease resistance. *Proc. Natl. Acad. Sci.* 115, 3174–3179. doi: 10.1073/pnas.1705927115

Zhu, J., Huang, F., Zhai, H., Zheng, Y., Yu, J., Chen, Z., et al. (2025). The Tetratricopeptide repeat protein TaTPR-B1 regulates spike compactness in bread wheat. *Plant Physiol.* 197, kiae546. doi: 10.1093/plphys/kiae546

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 17 July 2025
DOI 10.3389/fpls.2025.1589651

Check for updates

OPEN ACCESS

EDITED BY
Reyazul Rouf Mir,
Sher-e-Kashmir University of Agricultural Sciences and Technology, India

REVIEWED BY
Domancar Orona-Tamayo,
Centro de Innovación Aplicada en Tecnologías Competitivas (CIATEC), Mexico
Carmen Guerrero-Galán,
San Pablo CEU University, Spain

*CORRESPONDENCE
Si-Hai Wang
wangsh6688@163.com

RECEIVED 07 March 2025
ACCEPTED 23 June 2025
PUBLISHED 17 July 2025

CITATION
Wang S-H, Zhang C-G, Yang W, Chen J and Shi M (2025) Unveiling the mechanisms driving the rapid growth of *Malania oleifera* seedlings, a high-value root hemiparasitic plant. *Front. Plant Sci.* 16:1589651. doi: 10.3389/fpls.2025.1589651

COPYRIGHT
© 2025 Wang, Zhang, Yang, Chen and Shi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Unveiling the mechanisms driving the rapid growth of *Malania oleifera* seedlings, a high-value root hemiparasitic plant

Si-Hai Wang[1,2]*, Chuan-Guang Zhang[1,2], Wei Yang[1,2], Jian Chen[1,2] and Ming Shi[3]

[1]Yunnan Provincial Key Laboratory of Forest Plant Cultivation and Utilization, Yunnan Academy of Forestry and Grassland, Kunming, China, [2]Key Laboratory of the State Forestry Administration on Conservation of Rare, Endangered and Endemic Forest Plants, Kunming, China, [3]College of Forestry, Southwest Forestry University, Kunming, China

Certain root hemi-parasitic tree species hold significant economic value, yet they are challenging to cultivate artificially. Therefore, understanding how soil conditions and host plants influence the growth of these species is crucial. The endemic tree species *Malania oleifera*, native to the karst landscapes of southwest China, is highly valued for its seed oil, rich in nervonic acid. As a root hemiparasite, *M. oleifera* presents challenges for artificial cultivation, making it crucial to improve seedling survival and develop effective propagation methods for this and similar species. We used nutrient-rich and nutrient-deficient growth substrates, combined with four planting configurations involving host and non-host plants, to monitor the growth of *M. oleifera* seedlings. We then analyzed the transcriptomic differences between non-parasitic and parasitic plants that exhibited significant growth disparities. Vigorous host plants significantly enhance the growth of *M. oleifera* seedlings, while soil conditions exert a weaker influence. The host primarily promotes aboveground *M. oleifera* growth, with only limited impact on root development, resulting in an imbalance between the two. Endogenous hormone levels in the haustoria connected to the host exhibit substantial changes, with notable upregulation of genes related to hormone metabolism, stress responses, and antibiotic biosynthesis. Furthermore, the roots of host-associated *M. oleifera* seedlings show heightened responses to both biotic and abiotic stresses, along with key metabolic processes. An appropriate host enhances the overall adaptability, nutrient synthesis, and stress resistance of *M. oleifera* seedlings, all of which are essential for their growth, development, and survival.

KEYWORDS

haustorium, host plant, biomass, phytohormone, transcriptome analysis, soil conditions

1 Introduction

Root hemiparasitic plants have a wide range of uses, including as food, timber, and medicinal resources (Bell and Adams, 2011; Pignone and Hammer, 2016). However, because they rely on host plants for survival, cultivating them artificially presents some technical challenges (Pignone and Hammer, 2016; ArunKumar et al., 2022). Investigating the cultivation techniques for each economically valuable root hemiparasitic plant is both a fascinating and practically important endeavor. *Malania oleifera*, a monotypic species in the Olacaceae family, is an endangered tree endemic to China and native to the karst landscapes of southeastern Yunnan and western Guangxi provinces. The species was described scientifically in 1980 (Lee, 1980). In 1981, researchers discovered that the seed oil of *M. oleifera* is rich in nervonic acid (15-tetracosenic acid) (Ou, 1981). Nervonic acid plays a crucial role in human health, contributing to brain development and maintenance, improving memory, and delaying brain aging (Li Q. et al., 2019). The oil content of *M. oleifera* seed ranges between 48.3% and 67.9%, with nervonic acid comprising between 55.7% and 67.0% of the total fatty acids, making this species a valuable source of this nutrient (Ou, 1981; Tang et al., 2013; Wang et al., 2021). *M. oleifera* therefore stands out as a promising candidate for the discovery and development of nervonic acid resources. However, due to the scarcity of wild *M. oleifera* resources, cultivation would be the necessary path for its sustainable development and utilization.

Between the 1980s to about 2008, attempts to cultivate *M. oleifera* in the wild within its native habitat have yielded poor results (Liu, 1984; Huang et al., 2008), and the plants exhibited significant variation in individual growth. Some perished rapidly after a brief period of development, while others either grew slowly or not at all, and were susceptible to root diseases. Few individuals displayed robust growths, and overall, the survival rate was low, which was puzzling and which differed from the typical responses of other cultivated plants. Subsequently, researchers conducted afforestation experiments in four types of sites: open forests, shrubland, grasslands, and bare land. They found that the survival rates and growth of *M. oleifera* seedlings were highest in open forests, followed by shrubland, grasslands, and finally bare land. This phenomenon was attributed to the differences in growth environments caused by varying vegetation types (Lv et al., 2016). The unusual and seemingly unpredictable growth of *M. oleifera* after planting made effective cultivation challenging, and although the species has great economic value, this technical bottleneck in its development and utilization lies in the challenges of successful artificial cultivation.

By 2019, researchers had discovered that *M. oleifera* is a root hemiparasitic plant, a finding that offered new insights for the exploration of novel cultivation methods for the species (Li Y.P. et al., 2019, Li A.R. et al., 2019). Subsequent experiments were conducted to pair *M. oleifera* seedlings with various host plant species. While the seedlings successfully established parasitic relationships with the hosts, many of these host plants did not significantly enhance the growth of the seedlings (Li Y.P. et al., 2019; Chen et al., 2022). This outcome contrasts with the typical behavior observed in other root hemiparasitic plants, which often experience substantial growth benefits from their hosts (Chen et al., 2020; Thyroff et al., 2023). The specific conditions under which host plants can effectively promote the growth of *M. oleifera* seedlings therefore remain unclear.

In the absence of a host, the length of the normal growth period for root hemiparasitic plants depends on the characteristics of the plant itself and is also influenced by the growth environment. In the absence of a host, different soil nutrient levels have varying effects on the growth of different root hemiparasitic plant species (Kokla et al., 2022; Speetjens and Jacobs, 2023). When nutrients are artificially supplied, some root hemiparasitic plants can complete their entire life cycle in the absence of a host (Seel et al., 1993; Li et al., 2013). However, in some plants, seedling growth stagnates without a host even if nutrients are artificially supplied (Musselman, 1969; Těšitel et al., 2010). For *M. oleifera*, which is naturally distributed in karst regions, the extent to which soil conditions affect seedling growth it is still unknown.

The physiology of root hemiparasitic plants undergoes significant changes before and after a parasitic relationship with their hosts is established. Plant hormones play a crucial role in haustorium development, with changes in endogenous hormone levels contributing to this process. Through the haustorium, various substances, including secondary metabolites, RNAs, proteins, and nutrients, are transferred between parasitic plants and their hosts. These physiological changes are closely linked to the regulation of specific genes (Zhang et al., 2012; Shen et al., 2023; Ashapkin et al., 2023).

Transcriptomic analyses using RNA sequencing (RNA-seq) and *de novo* assembly have revealed the conservation of chlorophyll synthesis in root-parasitic Orobanchaceae (Wickett et al., 2011), as well as host-specific patterns of parasite gene expression at the interface between *Triphysaria versicolor* and its hosts (Honaas et al., 2013). The expression of haustorial genes differs in some hemiparasitic plants before and after establishing a parasitic relationship with the host (Zhang et al., 2015) as do levels of endogenous hormones levels and gene expression in the roots (Chen et al., 2021). Whole genome sequencing provides comprehensive background information for transcriptome sequencing, significantly enhancing the interpretation of transcriptomic data and enabling a more accurate understanding of the biological significance within the data. High-quality genomic data for *M. oleifera* have recently become available, and will enhance the analysis of its transcriptome data and biological functions (Yang et al., 2023). However, to date, a comprehensive understanding of the difference in gene expression differences in haustoria and roots is lacking, as is understanding of the intrinsic systemic connections between these differences and physiological changes. Moreover, the correlation between the changes at the micro level (gene expression differences and physiological changes) and the growth performance of the plants at the macro level needs further research.

Our preliminary experiments demonstrated that *M. oleifera* seedlings can establish a parasitic relationship with a fast-growing

subshrub, *Tithonia diversifolia*. This host plant effectively promotes the growth of *M. oleifera* seedlings. Based on these findings, we will address the following questions: (1) Which factor is more effective in promoting the growth of *M. oleifera* seedlings: the host plant or soil conditions? (2) Is there a direct correlation between the biomass of the host plant and the biomass of attached *M. oleifera* seedlings? (3) What are the differences in gene expression in haustoria and roots when plants are grown under non-parasitic and parasitic conditions, and how do these physiological changes affect seedling growth? This knowledge will not only enhance our understanding of the growth patterns of root hemiparasitic tree species, but also aid in developing effective artificial cultivation methods for economically valuable but rare root hemiparasitic trees.

2 Materials and methods

2.1 Plant materials

In October 2021, we collected a substantial number of mature fruits from wild *M. oleifera* plants in Guangnan County, Yunnan. We removed the peel and cleaned the seeds, we then selected seeds of uniform size and sowed them in a sand bed at the Kunming Arboretum. In early July 2022, once the seeds in the sand bed had germinated and developed 5–6 true leaves, we transplanted them into planting bags measuring 12 cm in diameter and 32 cm in height. The seedlings were then transplanted into one of two different substrates: either a nutrient-rich substrate (Substrate I) consisting of a 1:1 ratio of soil and organic fertilizer (V/V), or a nutrient-poor substrate (Substrate II) consisting of a 1:1 ratio of fine sand and perlite (V/V). The nutrient content of each substrate type is presented in Table 1. Seedling on each substrate type were either given access to a host plant, or not. The four resulting treatments were named as follows: substrate I without a host: S1N, substrate I with a host: S1H, substrate II without a host: S2N, and substrate II with a host: S2H. Each treatment was replicated 30 times, yielding a total of 120 planting bags. *M. oleifera* seedlings were planted at the center of each bag. In the planting bags with host plants, a *T. diversifolia* stem cutting, measuring 25 cm in length and 1.5 cm in diameter, was inserted obliquely beneath the root of the *M. oleifera* seedling. Two-thirds of the *T. diversifolia* cutting was buried in the substrate, maintaining a distance of 3 cm between the cutting and the seedling stem on the surface of the substrate. Preliminary experiments showed that stem cuttings of *T. diversifolia* root easily and tend to survive. The plants were planted in July, and the stem cuttings of *T. diversifolia* took root in about 10 days; The *M. oleifera* seedlings had established a parasitic relationship with their hosts after approximately one month. These planting bags were placed in a greenhouse with 50% shading and allowed to grow for one year. The average temperature in the greenhouse was 24°C in July and 10°C in January, and had a relative humidity of between 60% and 80%. These greenhouse conditions closely resemble the climatic conditions of the natural distribution area of *M. oleifera*. The planting bags were arranged with each bag spaced 30 cm apart. During the one-year experiment, hand weeding was performed as needed, and watering was carried out appropriately based on the moisture conditions of the substrate throughout the different seasons.

2.2 Monitoring growth data, sampling and harvesting

In early August and early October of 2022, as well as early January, April, June, and August of 2023, we measured the basal stem diameter, height, and leaf number of each *M. oleifera* seedling. Following the final growth measurements, the treatments S1N and S1H, that used the same substrate but exhibited significant differences in seedling growth, were selected for transcriptome and hormone analysis. Four types of samples were collected: roots from *M. oleifera* seedlings grown without a host (S1N_R) and their haustoria (S1N_H, as *M. oleifera* can form haustoria even in the absence of a host); as well as roots from seedlings grown with a host (S1H_R) and haustoria attached to the host (S1H_H). Each sample included three replicates, with each replicate consisting of 3 randomly selected individual roots or haustoria. All samples were immediately frozen in liquid nitrogen and stored at -80°C for future analysis. Root samples were used only for RNA sequencing, while haustorium samples were also used for hormone analysis.

In order to monitor the growth of the host and parasite plants, we next rinsed the substrate in each planting bag thoroughly with clean water. In bags containing host plants, we counted the number of haustoria formed by each *M. oleifera* seedling parasitizing the roots of *T. diversifolia*. The seedlings were then carefully separated from the host roots, ensuring that the root systems of the *M. oleifera* seedlings remained intact. We then used an electronic balance to measure the fresh weight of the aboveground biomass and the root biomass of each *M. oleifera* seedling, along with the fresh weight of the aboveground biomass of *T. diversifolia*. The biomass samples were placed in an oven at 60°C for 48 hours, after which we measured the dry weight of each sample.

TABLE 1 Nutrient content of the two planting substrates.

Substrate type	SOM (g/kg)	TN (g/kg)	TP (g/kg)	TK (g/kg)	HN (mg/kg)	AP (mg/kg)	AK (mg/kg)
Substrate I	132.50 ± 2.63	5.31 ± 0.14	2.52 ± 0.23	6.85 ± 0.10	340.17 ± 11.50	44.22 ± 8.51	251.17 ± 4.56
Substrate II	1.86 ± 0.57	0.07 ± 0.01	0.07 ± 0.02	2.50 ± 0.96	4.00 ± 0.89	1.47 ± 0.68	13.17 ± 2.14

After thoroughly mixing each substrate, six samples were randomly selected for measurement. SOM, soil organic matter; TN, total N; TP, total P; TK, total K; HN, hydrolysable N; AP, available P; AK, available K.

2.3 RNA extraction and sequencing, and endogenous hormone measurement in haustoria

Total RNA was extracted from tissue samples using TRIzol Reagent (Ambion, Cat# 15596018) following the manufacturer's protocol (For detailed methods, see Supplementary Data Sheet 1). Concentration and purity were measured using a NanoDrop spectrophotometer (Thermo Scientific NanoDrop 2000, Thermo Scientific, Waltham, Massachusetts, USA). RNA integrity was evaluated using RNA-specific agarose gel electrophoresis and the Agilent 2100 Bioanalyzer with the RNA 6000 Nano Kit (5067-1511, Agilent Technologies Inc., California, USA). Poly-A tail-containing mRNA was enriched from total RNA using Oligo(dT) magnetic beads, and the RNA was then sheared into fragments of approximately 300 bp using ion fragmentation. Using the RNA as a template, the first strand of cDNA was synthesized with random hexamer primers and reverse transcriptase, followed by the synthesis of the second cDNA strand using the first strand as a template. After library construction, PCR amplification was used to enrich the library fragments, and library quality was assessed with the Agilent 2100 Bioanalyzer. Paired-end (PE) sequencing was subsequently performed on the libraries using next-generation sequencing (NGS) technology on the Illumina platform. The raw sequencing data was filtered to generate high-quality clean data, which were then aligned to the reference genome of *M. oleifera* (Yang et al., 2023). Gene expression levels were calculated based on the alignment results, followed by differential expression analysis, enrichment analysis, and clustering analysis of the samples.

Samples were prepared for LC-MS/MS analysis as follows. 40 mg of lyophilized sample was weighed into a 2 mL brown centrifuge tube, 1 mL each of methanol and mixed internal standard stock solution was accurately measured and added to the sample. Samples were sonicated for 10 min, then transferred to a metal bath and shaken. After 4 hours, samples were centrifuged at 12000 rpm for 10 min at 4 °C, and the entire supernatant was removed. 0.5 mL methanol was added to the remaining residue and samples were further shaken for 2 h in a metal bath. The extracts were then centrifuged through a 0.22 μm filter membrane and placed in the injection vial for LC-MS/MS analysis (Li et al., 2016).

2.4 qRT-PCR verification of DEGS

A total of 16 differentially expressed genes were selected for expression analysis, including 6 genes specifically identified from haustoria, 7 genes specifically identified from roots, and 3 genes selected from both haustoria and roots. RNA extraction and complementary DNA (cDNA) synthesis were performed following the protocols described in Section 2.3. The forward and reverse primers were designed by Primer Premier (version 6.0, Premier Biosoft Inc., CA, USA). The reaction was performed using the Power qPCR PreMix (Genecopoeia) in a 96-Well PCR plate (PCR-96-FLT-C, Axygen). The reaction conditions were established according to the manufacturer's instructions, with each reaction performed in triplicate. The *actb* was employed as the reference gene. The $2^{-\Delta\Delta Ct}$ method was employed to calculate relative expression levels. The detailed experimental procedures and specific primers are listed in Supplementary Data Sheet 1.

2.5 Statistical analysis

The cumulative growth from each monitoring event and the harvested biomass of *M. oleifera* seedlings were analyzed using ANOVA (One-way analysis of variance). The annual cumulative growth of each *M. oleifera* seedling, the harvested biomass, the number of haustoria connected to the host, and the aboveground biomass of *T. diversifolia* (only for plants with a host) were analyzed for correlations using Pearson correlation analysis. The hormone content in the parasitic haustoria was analyzed using an independent samples t-test. All statistical analyses were performed using the R statistical package.

We performed statistical analysis on the raw data (Raw Data) of each sample, including sample name, Q30, percentage of ambiguous bases, and Q20(%) and Q30(%). Sequencing data, including adapter-contaminated and low-quality reads, were then filtered using Fastp to remove 3' adapter sequences and reads with an average quality score below Q20 were discarded. The filtered reads were aligned to the reference genome (GCF_029873635.1_ASM2987363v1_genomic.fna) using the HISAT2 software and the mapping information was calculated.

We used HTSeq to count the number of reads aligned to each gene, which served as the raw gene expression values. FPKM (Fragments Per Kilo bases per Million fragments) was used to normalize the expression levels, and genes with FPKM > 1 were considered to be expressed. We performed differential expression analysis using DESeq, with the criteria for selecting differentially expressed genes set as |log2FoldChange| > 1 and a significance threshold of P-value < 0.05. Differentially expressed genes (DEGs) were analyzed separately using Gene Ontology (GO) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) database, and were annotated and enriched to obtain functional analyses and pathway results for DEGs.

3 Results

3.1 The growth dynamics of *M. oleifera* under different planting scenarios

The cumulative growth of the monitored growth indicators (increase in stem diameter, plant height, and leaf number) showed similar growth trends across the four planting methods (Figures 1A-C). The fastest-growing plants were all in treatment S1H, and showed significant growth differences as early as April of the second year. The increases in stem diameter, plant height, and leaf number were all significantly greater than those of plants subjected to the other three treatments (P<0.01). By August of the second year, the growth differences had become more pronounced,

10.3389/fpls.2025.1589651

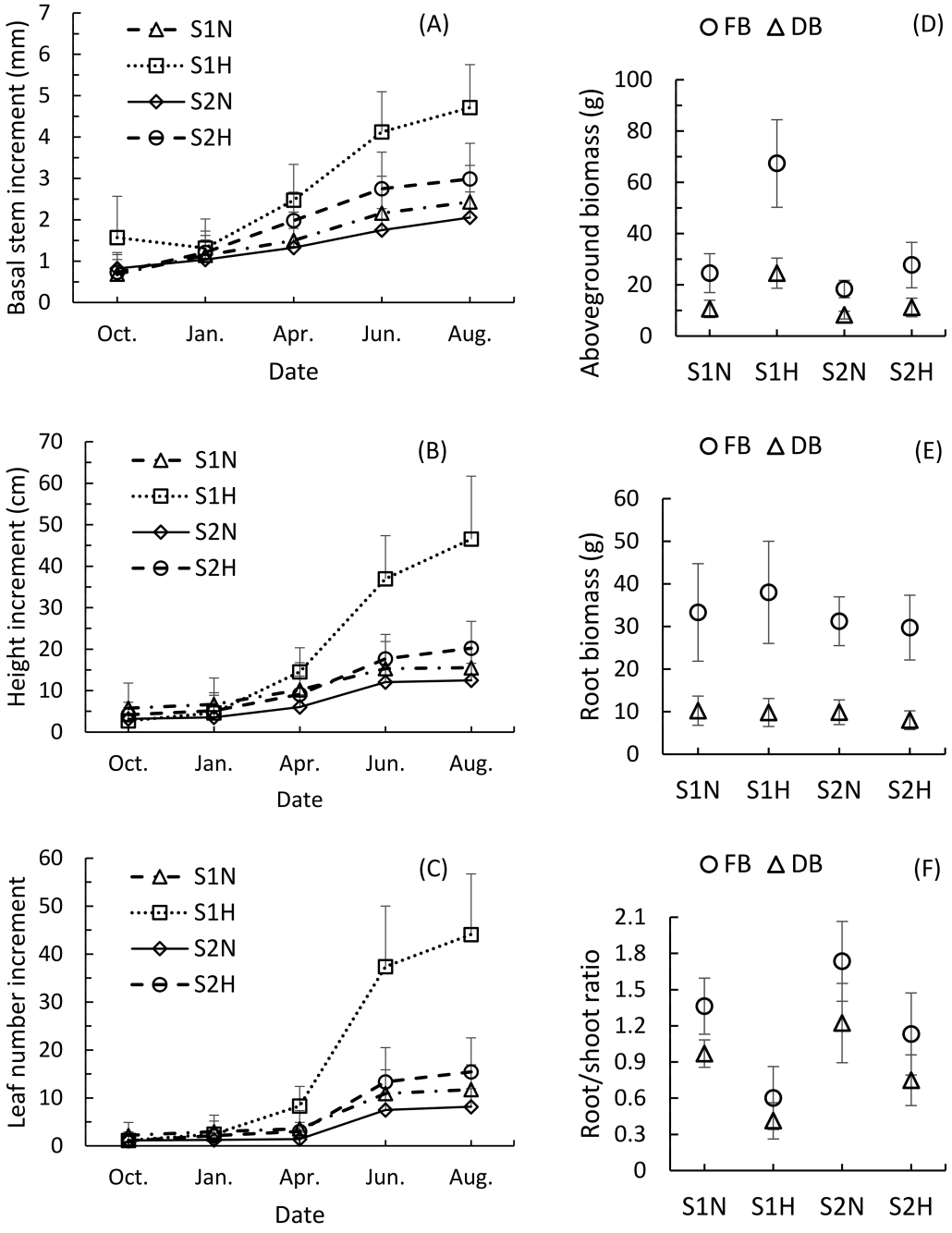


FIGURE 1
Comparison of the growth dynamics and biomass of *Malania oleifera* seedlings under four planting treatments (mean ± SD, n=30). **(A)** Annual growth dynamics of basal stem diameter. **(B)** Annual growth dynamics of plant height. **(C)** Annual growth dynamics of plant leaf number. **(D)** Aboveground fresh and dry biomass of plants for a whole year. **(E)** Root fresh biomass and dry biomass of plants for a whole year. **(F)** Annual root-to-shoot ratio analysis of plant growth. FB, Fresh biomass; DB, Dry biomass. S1N: nutrient-rich substrate without host; S1H, nutrient-rich substrate with host; S2N, nutrient poor substrate without host; S2H, nutrient-poor substrate with host.

with S1H plants showing highly significant differences in growth compared to plants subjected to the other three treatments (S1N, S2N, S2H) ($P<0.001$).

The second fastest-growing group was S2H. By August of the second year, the increases in stem diameter, plant height, and leaf number of plants in treatment group S2H were significantly greater than those of plants subjected to treatments S1N ($P<0.05$) and S2N ($P<0.01$). S1N plants grew faster than those in the S2N group, but by August of the second year, the differences in these indicators were not significant ($P>0.05$). These findings indicate that selecting an effective host plant is conducive to promoting the growth of *M. oleifera* seedlings, and, while favorable soil conditions can also directly enhance seedling growth, though the effect is not significant.

The fresh and dry weights of the aboveground biomass of *M. oleifera* seedlings across the four treatments showed similar trends to their growth patterns (Figure 1D). S1H plants had the highest aboveground fresh and dry biomass (67.32 ± 17.09g, 24.53 ± 5.88g), followed by plants in groups S2H (27.68 ± 8.87g, 11.14 ± 3.63g), S1N (24.55 ± 7.59g, 10.59 ± 3.38g), and S2N (18.33 ± 3.39g, 8.16 ± 1.51g). The fresh and dry aboveground biomass of plants under treatment S1H was significantly different from plants under the other three treatments ($P<0.001$). The fresh and dry aboveground biomass of S2H plants were significantly greater than those of S2N plants ($P<0.01$), but the difference between S2H and S1N plants was not significant ($P>0.05$). The differences in fresh and dry aboveground biomass between plants subjected to treatments S1N and S2N were significant ($P<0.05$).

The root biomass showed a different pattern from the aboveground biomass (Figure 1E). Plants in group S1H had the highest fresh root weight (38.01 ± 12.00g), followed by those in S1N (33.29 ± 11.45g), S2N (31.23 ± 5.73g), and S2H (29.73 ± 7.60g). Plants subjected to treatment S1N had the highest dry root weight (10.23 ± 3.43g), followed by those in treatment groups S2N (9.86 ± 2.90g), S1H (9.80 ± 3.29g), and S2H (7.99 ± 2.20g). Significant differences in fresh root biomass were only found between S1H and S2N, and S1H and S2H ($P<0.01$), while no significant differences were observed between any other pairs ($P>0.05$). The only significant differences in dry root biomass were found between plants in the S2H treatment group and those subjected to the other three treatments ($P<0.05$).

However, the root-to-shoot ratio exhibited the opposite trend compared to aboveground growth (Figure 1F). Plants subjected to treatment S1H had the smallest root-to-shoot ratio, with fresh and dry values of 0.60 ± 0.26 and 0.41 ± 0.15, respectively. In comparison, S2H plants had root-to-shoot ratios of 1.13 ± 0.34 and 0.75 ± 0.21, S1N plants had 1.36 ± 0.23 and 0.97 ± 0.11, and S2N had 1.73 ± 0.33 and 1.22 ± 0.33. The aboveground biomass growth of *M. oleifera* seedlings parasitizing *T. diversifolia* increased significantly compared with that of seedlings without a host plant, while root biomass growth lagged, resulting in an imbalance between aboveground and belowground development.

Under different soil conditions, the growth of the host plant *T. diversifolia* showed highly significant differences ($P<0.01$). The aboveground biomass of *T. diversifolia* in the nutrient-rich S1H treatment group had fresh and dry weights of 142.46 ± 76.46g and 17.43 ± 8.85g, respectively, while those in the nutrient-poor S2H treatment group had fresh and dry weights of 18.61 ± 2.20g and 2.44 ± 2.87g, respectively (Figure 2A). The number of haustoria of *M. oleifera* parasitizing the roots of *T. diversifolia* also showed highly significant differences ($P<0.01$) across different substrates, with 26.90 ± 10.80 haustoria produced by plants subjected to the S1H treatment, but only 11.83 ± 7.18 in the S2H treatment group (Figure 2B). Further correlation analysis revealed that the aboveground biomass (fresh and dry weight) of *T. diversifolia* was significantly positively correlated with *M. oleifera* basal stem increment ($P<0.05$), height increment ($P<0.01$), leaf number increment ($P<0.01$), fresh and dry aboveground biomass ($P<0.01$), and the number of parasitic haustoria ($P<0.01$), but showed no correlation with its root biomass (fresh and dry weight). Additionally, the number of haustoria was highly significantly positively correlated ($P<0.01$) with basal stem increment, height increment, leaf number increment, and the fresh and dry aboveground biomass of *M. oleifera* ($P<0.01$) (Figure 2C). Clearly, the growth of the host plant *T. diversifolia* significantly promoted the aboveground biomass growth of *M. oleifera*, while having little effect on root growth.

3.2 Quality assessment of transcriptome sequencing data from haustoria and roots

82.2 G of raw data were obtained from transcriptome sequencing of twelve samples (Supplementary Table S1). The raw data for all twelve samples had a Q20 quality score of 98.05% and a Q30 quality score of 94.28%. The GC content of the raw reads ranged from 43.39% to 46.22%. The clean data mapping rates to the *M. oleifera* genome were more than 93.66%, except for sample S1H-R06 (73.56%). The haustorium and root transcriptome sequencing data with high quality and mapping ratios were used for subsequent RAN analysis. The correlation analysis of gene expression levels between biological replicates showed that the Pearson correlation coefficients were all greater than 0.89, except for the value (0.77) between sample S1N_H01 and S1N_H03 (Supplementary Table S2).

3.3 Differential gene expression in *M. oleifera* haustoria and roots under non-parasitic and parasitic conditions

17023 (S1N_H), 16963 (S1N_R), 16862 (S1H_H), and 16759 (S1H_R) genes (expressed in all three replicates) were expressed in the haustoria and roots of plants in the S1N and S1H treatment groups, respectively (Figure 3A). The numbers of DEGs for S1N_H vs S1H_H, S1N_R vs S1H_R, S1N_R vs S1N_H and S1H_R vs S1H_H were 1087, 1094, 1661 and 2086, respectively (Figure 3B). There were 591, 447, 681 and 1001 up-regulated DEGs between S1N_H vs S1H_H, S1N_R vs S1H_R, S1N_R vs S1N_H and S1H_R vs S1H_H, respectively, and 496, 647, 980 and 1085 down-regulated DEGs, respectively (Figure 3D). The cluster analysis demonstrated that the differences in gene expression were significant among these four treatment groups (Figure 3C).

3.4 Functional changes in the haustoria of *M. oleifera* seedlings between non-parasitic and parasitic states

GO annotations and KEGG pathways analyses were used to analyze changes in haustorium function in non-parasitic and parasitic *M. oleifera* plants. The number of GO terms that were significantly enriched in up-regulated DEGs in the haustoria of *M. oleifera* plants without versus with a host (S1N_H vs S1H_H), was

10.3389/fpls.2025.1589651

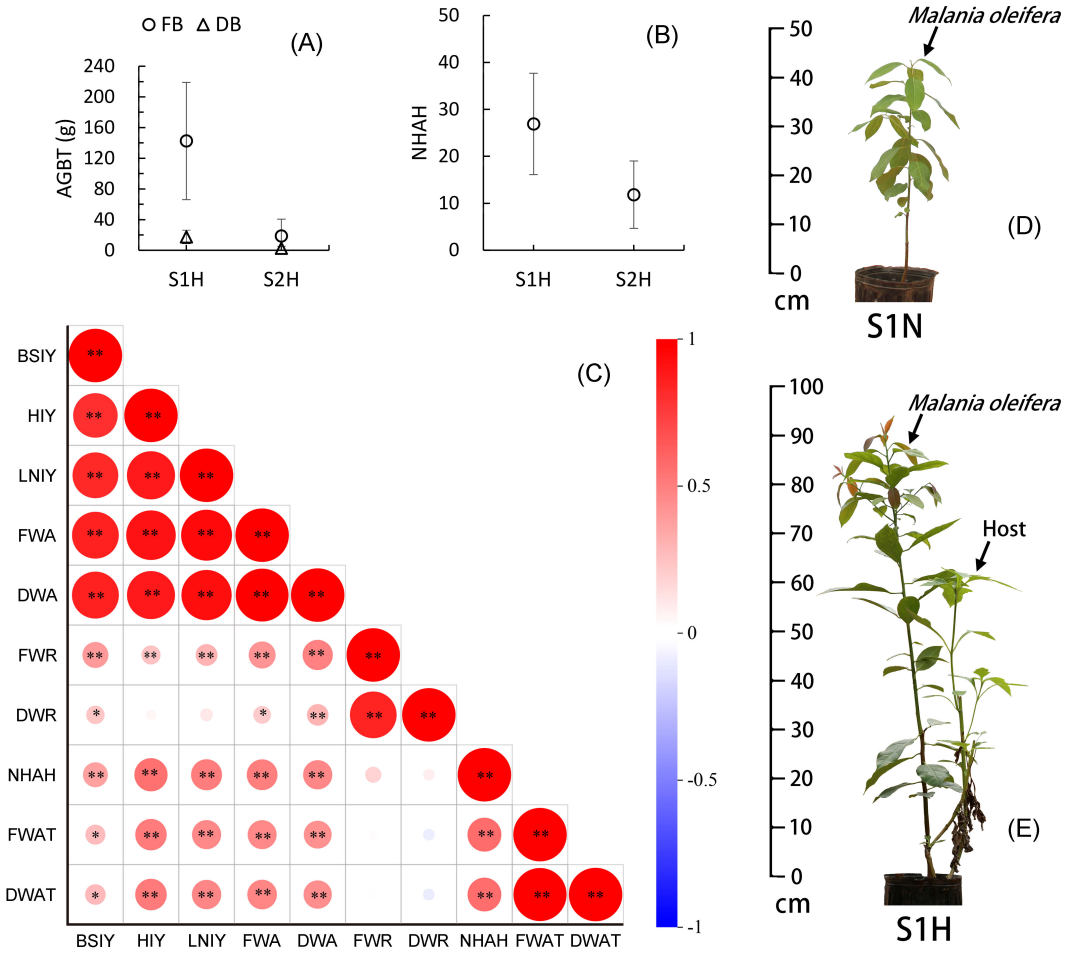


FIGURE 2
Correlation between *Malania oleifera* seedlings and host growth. **(A)** Aboveground biomass of *T. diversifolia* (AGBT) (mean ± SD, n=30). **(B)** the number of *M. oleifera* haustoria attached to the host (NHAH) (mean ± SD, n=30). **(C)** AGBT, NHAH and Seedling Growth Correlations in *M. oleifera* (n=60). **(D, E)** Growth comparison of representative plant under S1N and S1H treatments (To better display the parasitic plant, some branches of the host were removed). FB, Fresh biomass; DB, Dry biomass; BSIY, Basal stem increment for a whole year; HIY, Height increment for a whole year; LNIY, Leaf number increment for a whole year; FWA, Fresh weight of aboveground biomass; DWA, Dry weight of aboveground biomass; FWR, Fresh weight of root biomass; DWR, Dry weights of root biomass; FWAT, Fresh weight of aboveground biomass of *T. diversifolia*; DWAT, Dry weight of aboveground biomass of *T. diversifolia*. *Significance at $P < 0.05$, **significance at $P < 0.01$.

much higher than in down-regulated DEGs. A total of 140 GO terms were associated with up-regulated DEGs (FDR < 0.01), while only 12 GO terms were linked to down-regulated DEGs. Most GO terms for up-regulated DEGs were related to biological processes, including the regulation of hormone metabolism and biosynthesis (e.g., jasmonic acid and salicylic acid), response to ozone, antibiotic biosynthesis, and lipid metabolism and biosynthesis (Supplementary Table S3). In contrast, the down-regulated DEGs were mainly enriched in functions associated with the cell periphery, cell wall, and external encapsulating structure.

The KEGG pathways identified in the DEGs in the haustoria of *M. oleifera* plants without versus with a host (S1N-H vs S1H-H) were mainly associated with the biosynthesis of other secondary metabolites, signal transduction, carbohydrate metabolism, metabolism of terpenoids and polyketides, metabolism of other amino acids, lipid metabolism and energy metabolism (Figure 4C). The key functions of these identified up-regulated DEGs were enriched in the pathways of Zeatin biosynthesis, plant hormone signal transduction, carotenoid biosynthesis, photosynthesis-antenna proteins and plant-pathogen interactions (Figure 4A). The down-regulated DEGs were mainly enriched in the pathways phenylpropanoid biosynthesis, starch and sucrose metabolism, pentose and glucuronate interconversion, plant hormone signal transduction and cyanoamino acid metabolism (Figure 4B).

The results of the GO and KEGG analyses suggest that when the *M. oleifera* haustorium is connected to a host plant, resistance and adaptability of the hemiparasite are enhanced through regulation of hormone signaling, secondary metabolism, and activation of defense-related pathways. This regulatory mechanism involves the metabolism of various hormones (such as jasmonic acid and salicylic acid), the synthesis of secondary metabolites, protection against antioxidants, and adjustments in cell structure, enabling the plant to effectively respond to a range of biotic and abiotic stresses from the external environment.

10.3389/fpls.2025.1589651

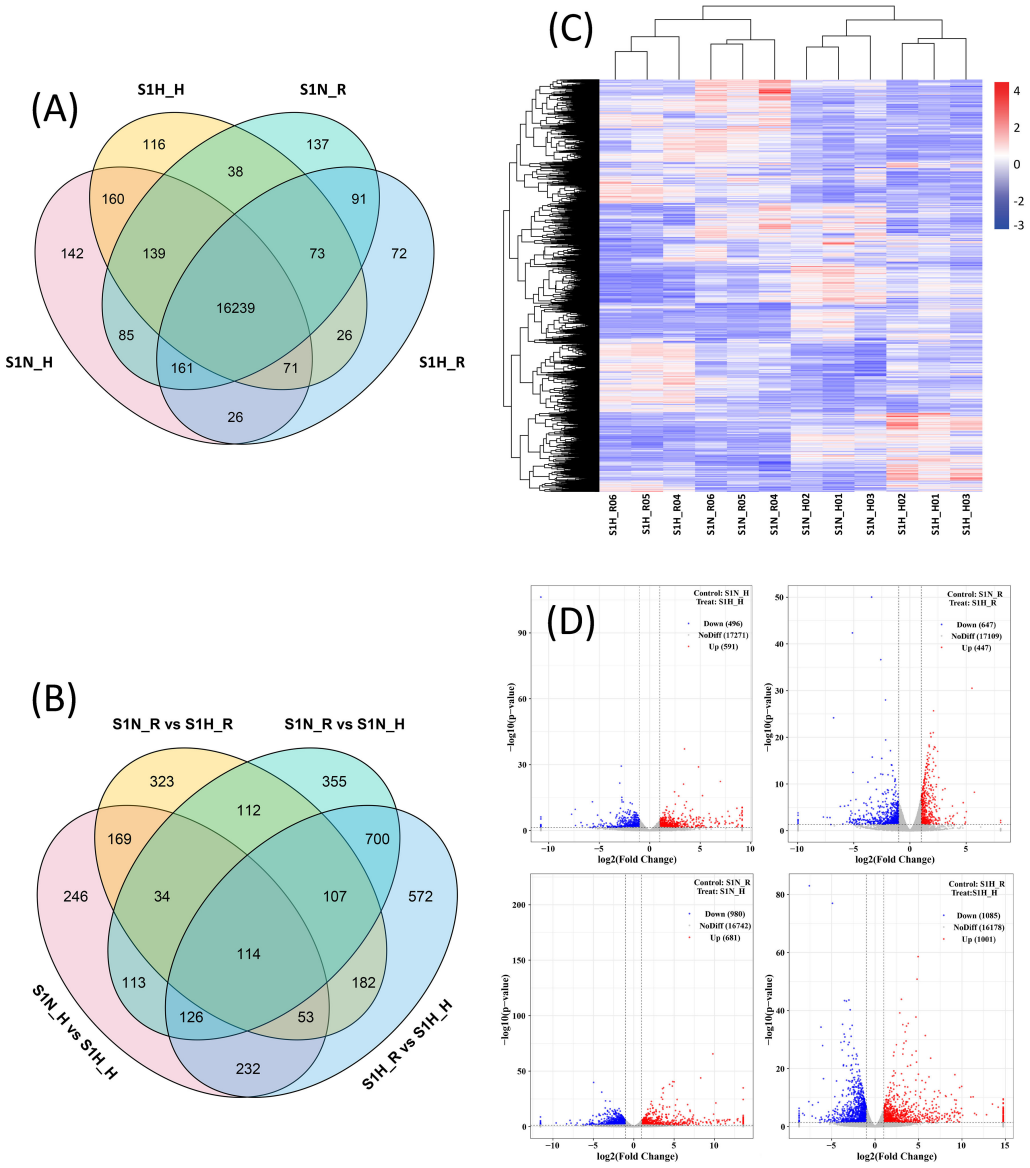


FIGURE 3
Gene expression in haustoria and roots of *Malania oleifera* under different planting treatments. **(A)** Number of expressed genes in *M. oleifera* haustoria and roots. **(B)** Number of DEGs. **(C)** The DEGs cluster heat map. **(D)** Volcano plot of DEGs.

3.5 Functional changes in the roots of *M. oleifera* seedlings in non-parasitic and parasitic states

GO annotations and KEGG pathways analyses were also used to analyze changes in root function between plants in non-parasitic and parasitic states. Eighty-two GO terms were associated with up-regulated DEGs (FDR < 0.01), and 42 GO terms with down-regulated DEGs (Supplementary Table S3), in *M. oleifera* roots under both non-parasitic and parasitic conditions (S1N_R vs S1H_R). Similar to haustorial DEGs, the GO terms related to up-regulated DEGs in roots were primarily involved in biological processes. The key functions of these up-regulated DEGs were enriched in responses to chitin, organonitrogen and oxygen-containing compounds, drug, chemical, other organisms, internal and external biotic stimuli, organic substances, stress, as well as in transcription regulator activity and DNA-binding transcription factor activity. In contrast, the down-regulated DEGs were mainly enriched in functions related to the cell periphery, cell wall, external encapsulating structure, enzymatic activity (including oxidoreductase, transferase, glucosyltransferase and quercetin glucosyltransferase), saponin metabolism and biosynthesis, responses to starvation, water and fluid transport.

The KEGG pathways for S1N-R vs S1H-R DEGs of roots were mainly associated with biosynthesis of other secondary metabolites, carbohydrate metabolism, signal transduction, energy metabolism and metabolism of terpenoids and polyketides (Figure 4F). The key functions of these up-regulated DEGs in S1N-R vs S1H-R were

10.3389/fpls.2025.1589651

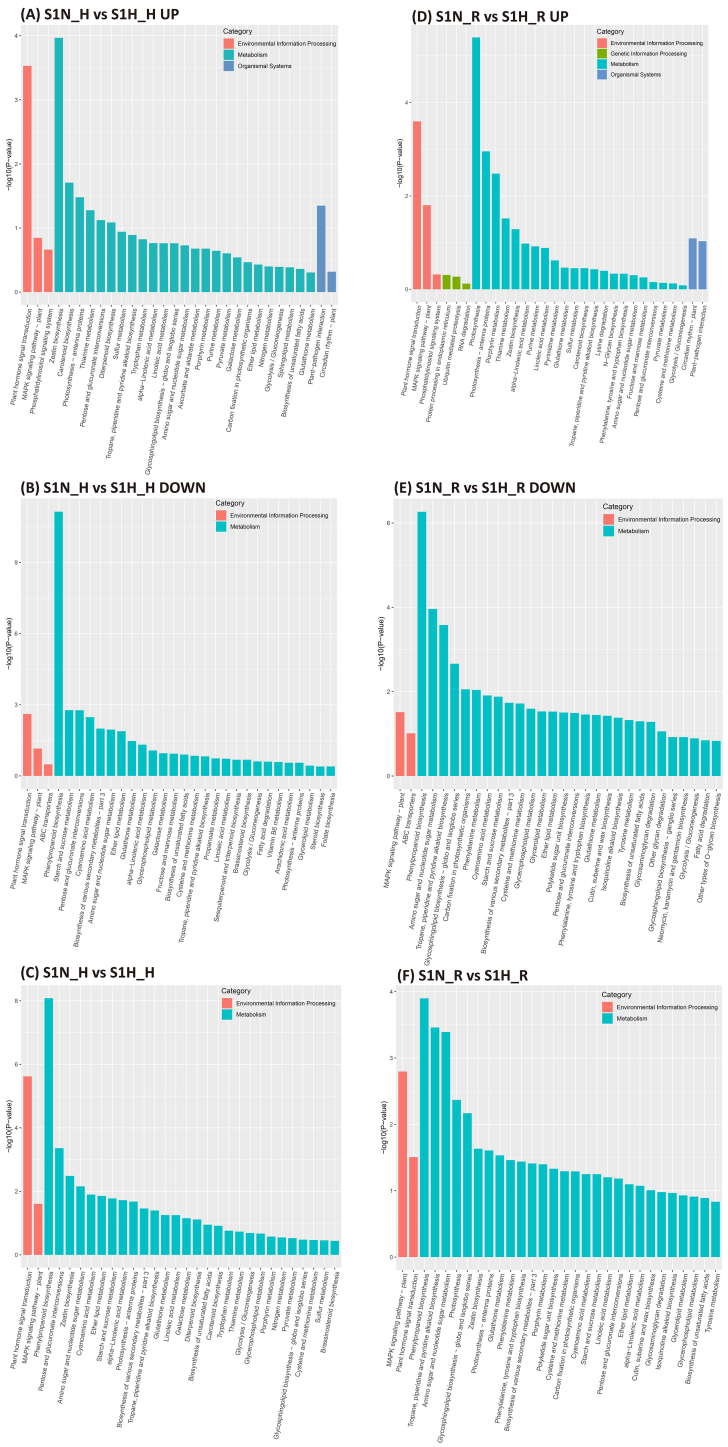


FIGURE 4
KEGG enrichment analysis in *Malania oleifera* haustoria and roots in the absence and presence of the host plant. **(A)** Up-regulated genes in S1N_H vs S1H_H. **(B)** Down-regulated genes in S1N_H vs S1H_H. **(C)** DEGs in S1N_H vs S1H_H. **(D)** Up-regulated genes in S1N_R vs S1H_R. **(E)** Down-regulated genes in S1N_R vs S1H_R. **(F)** DEGs in S1N_R vs S1H_R. S1N, nutrient-rich substrate without host; S1H, nutrient-rich substrate with host.

enriched in the pathways of photosynthesis, plant hormone signal transduction, porphyrin metabolism, the MAPK signaling pathway and thiamine metabolism (Figure 4D), and the down-regulated DEGs were mainly enriched in the pathways of phenylpropanoid biosynthesis, amino sugar and nucleotide sugar metabolism, tropane piperidine and pyridine alkaloid biosynthesis, glycosphingolipid biosynthesis (Figure 4E).

The GO and KEGG analysis data collectively indicate that the roots of *M. oleifera* seedlings, when associated with a host, actively participate in responses to environmental stresses, pathogen

defense, biosynthesis of protective secondary metabolites, hormone-mediated signaling, and maintenance of cellular homeostasis. This highlights the complex regulatory networks in *M. oleifera* that balance growth, defense, and adaptation under stress conditions.

3.6 Functional comparison of the root systems of non-parasitic and parasitic *M. oleifera* seedlings

To investigate the impact of parasitism on the entire root system of *M. oleifera* seedlings, we compared the functional changes in DEGs (both in haustorial organs and roots) under non-parasitic conditions (S1N_R vs S1N_H) and parasitic conditions (S1H_R vs S1H_H). In the absence of parasitism, the DEGs between haustoria and roots (S1N_R vs S1N_H) were predominantly associated with basic metabolic and defense-related functions of the plant (Supplementary Table S4). These include processes such as cell wall biosynthesis and tissue development (e.g., GO:0005618 cell wall, GO:0009505 plant-type cell wall, GO:0071554 cell wall organization or biogenesis, GO:0099402 plant organ development, ko00940 phenylpropanoid biosynthesis, ko04075 plant hormone signal transduction), as well as hormone metabolism and fundamental responses to various stimuli under abiotic stress conditions (GO:0042445 hormone metabolic process, GO:0009725 response to hormone, GO:0010200 response to chitin, GO:0001101 response to acid chemical, GO:0050896 response to stimulus, ko04075 plant hormone signal transduction, ko00905 brassinosteroid biosynthesis, ko04016 MAPK signaling pathway), and core metabolic processes (GO:0005886 plasma membrane, GO:0006811 ion transport, ko00040 pentose and glucuronate interconversions, ko00500 starch and sucrose metabolism, ko00910 nitrogen metabolism, ko00270 cysteine and methionine metabolism, ko01040 biosynthesis of unsaturated fatty acids).

Under parasitic conditions, the DEGs between haustoria and roots (S1H_R vs. S1H_H) not only support basic growth functions but also specifically enhance defense responses and adaptation to biotic stress (Supplementary Table S5). This is achieved by regulating specific hormones (GO:0080140 regulation of jasmonic acid metabolic process, GO:0080141 regulation of jasmonic acid biosynthetic process, GO:0080142 regulation of salicylic acid biosynthetic process, GO:0010337 regulation of salicylic acid metabolic process, ko04075 plant hormone signal transduction) and by strengthening defense-related metabolic pathways (GO:0051707 response to other organisms, GO:0043207 response to external biotic stimulus, GO:0009607 response to biotic stimulus, ko00950 isoquinoline alkaloid biosynthesis, ko04626 plant-pathogen interaction) to counter insect or pathogen invasion. In summary, the root systems of *M. oleifera* seedlings without hosts sustains only the basic functions necessary for plant structure and growth. In contrast, in plants with a suitable host, the root systems significantly enhance the precise regulation of defense and stress responses, illustrating a more dynamic and adaptive response mechanism in plants under complex environmental conditions.

3.7 Changes of hormone content in haustoria and KEGG enrichment

We found significant differences in the hormone content of non-parasitic and parasitic *M. oleifera* haustoria (Supplementary Table S6). Of the 23 hormones detected in the haustoria, 6 showed significant differences ($P < 0.05$), while 3 displayed marginal differences ($P = 0.089$, $P = 0.072$, $P = 0.061$). In the parasitic haustoria, the contents of JA-Ile, JA, SAG, TZR, ICAId, ACC increased by 145.9 times, 16.3 times, 12.6 times, 5.6 times, 2.3 times and 1.3 times, respectively, while the contents of GA24, IAA and ABA decreased by 5.6 times, 4.2 times and 3.2 times, respectively. The levels of the 23 hormones detected in the haustoria showed varying degrees of correlation with the differential expression of 289 genes in the haustoria and 300 genes in the roots (Supplementary Figure S1). Metabolite pathway analysis showed that there were 15 hormones that affected KEGG pathways, with the main pathways affected being biosynthesis of plant secondary metabolites, biosynthesis of plant hormones, metabolic pathways, biosynthesis of secondary metabolites, plant hormone signal transduction, and zeatin biosynthesis (Figure 5). This suggests that during the development of *M. oleifera*, the hormone levels in haustoria parasitizing host plants undergo significant changes. These hormonal fluctuations influence various metabolic processes in the *M. oleifera*, ultimately impacting its growth.

3.8 Validation of DEGs using qRT-PCR

To ensure the accuracy and reliability of the sequencing data under both non-parasitic and parasitic growth conditions, we selected 16 differentially expressed genes (DEGs), including 6 related to haustoria, 7 related to roots, and 3 shared by both tissues, for expression validation through quantitative real-time PCR (qRT-PCR). As shown in Figure 6, the qRT-PCR and RNA-Seq results were consistent. These verified genes involve abscisic acid receptor PYL9 (LOC131166841), auxin response factor (LOC131165241), transcription factor MYC2-like (LOC131163893), WRKY transcription factor WRKY24-like (LOC131147577), protein TIFY 9 (LOC131144836), two-component response regulator (LOC131164606), WRKY transcription factor (LOC131166883), and calmodulin-like protein (LOC131158404), which are closely related to the changes in the main physiological activities of *M. oleifera* under parasitic conditions.

4 Discussion

4.1 The impact of the host on the growth of *M. oleifera* seedlings

In the early stages of growth, root hemiparasitic plants can maintain a normal growth period for some time by absorbing nutrients stored in their seeds or by directly absorbing nutrients from the soil (Musselman, 1969; Seel et al., 1993; Tesitel et al., 2010; Li et al., 2013). *M. oleifera* seeds are relatively large, with a diameter

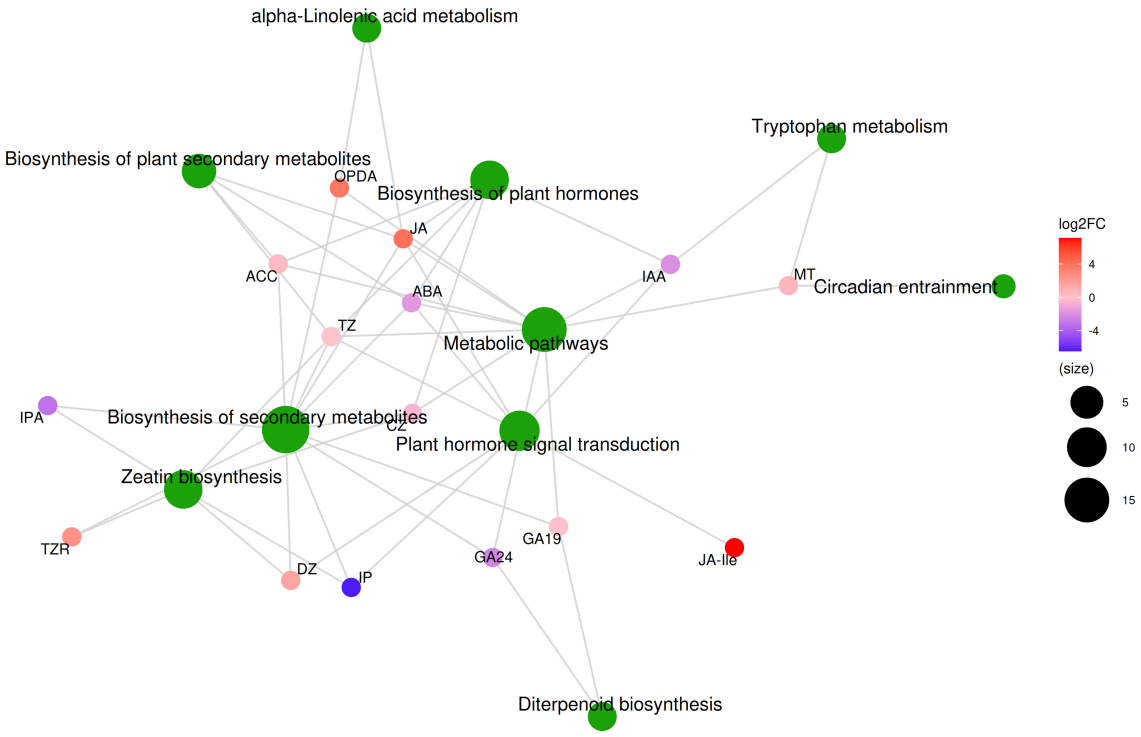


FIGURE 5
KEGG metabolic compound molecular network diagram.

of 2–5 cm, and are rich in oils (Wang et al., 2021), and after germination, these abundant nutrients are transferred to the seedling. Due to this nutrient reserve, seedlings grow vigorously in their first year, after which growth gradually weakens until it eventually stagnates. During this process, the concentrations of N, P, and K in most plant tissues, along with starch granule content, progressively decrease. Some individuals with weakened growth can survive for more than three years (Chen et al., 2024). Our study shows that *M. oleifera* seedlings grown in nutrient-rich substrate without a host (S1N treament) does not exhibit significant

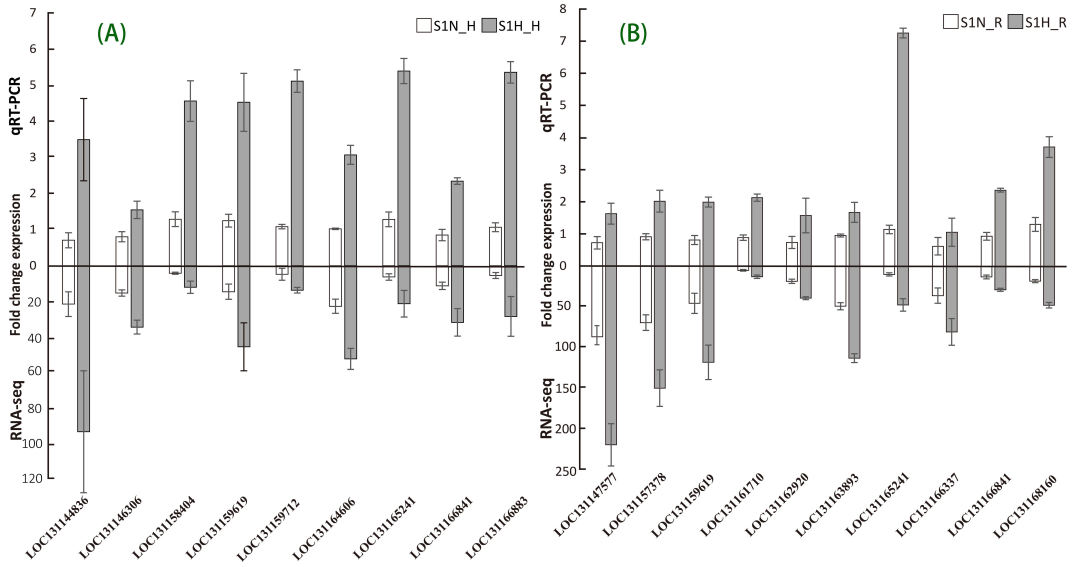


FIGURE 6
Comparison of the expressions levels of DEGs determined with RNA-Seq and qRT-PCR (Expression of the target DEGs was normalized to that of the β-actin gene). **(A)** Verification of DEGs in *M. oleifera* haustoria. **(B)** Verification of DEGs in *M. oleifera* roots.

differences in increments of basal stem diameter, height, or leaf number compared to plants grown in nutrient-poor substrate (S2N treatment) ($P > 0.05$). This suggests that favorable soil conditions alone have limited effects on the growth of *M. oleifera*. In the presence of a host, however, *M. oleifera* growth is significantly promoted. Plants cultivated under the S1H treatment (nutrient rich substrate and in the presence of a host) show highly significant differences in basal stem diameter, height, and leaf number compared to the other three treatments ($P < 0.001$). When a host is present, nutrient-poor substrates (S2H treatment) promote *M. oleifera* seedling growth more effectively than do nutrient-rich substrates in the absence of a host (S1N treatment). The two experimental substrates (Substrate I: soil + organic fertilizer; Substrate II: fine sand + perlite) differed significantly not only in nutrient content, but also potentially in texture, water-holding capacity, microbiota, and possibly other aspects. However, despite these substantial variations in edaphic properties, the growth parameters of *M. oleifera* seedlings cultivated without host plants showed no statistically significant differences between the two substrates ($P>0.05$ for S1N vs S2N treatments). This suggests that soil conditions alone have a limited impact on seedling development. Instead, these findings further emphasize the critical role of a host plant in promoting the normal growth of *M. oleifera* seedlings.

Faster-growing host plants are known to more effectively promote the growth of parasitic plants than slow-growing hosts (Hautier et al., 2010). In our study, the growth of *M. oleifera* seedlings was closely connected to the growth of the host plant. The aboveground biomass (fresh and dry weight) of the host *T. diversifolia* was highly significantly correlated with the height increase, leaf number, and aboveground biomass (fresh and dry weight) of *M. oleifera* seedlings ($P < 0.01$) and significantly correlated with basal stem diameter increase ($P < 0.05$; Figure 2C). This indicates that the rapid growth of *T. diversifolia* effectively promoted the rapid growth of *M. oleifera*. Sufficient nutrient acquisition from the host is essential to ensure the rapid growth of the parasitic plant. Previous studies have used tree seedlings of the species *Pinus armandii*, *Pistacia weinmannifolia*, and *Alnus ferdinandi-coburgii*, as well as herbaceous plants including *Solanum tuberosum*, *Chlorophytum comosum*, and *Artemisia argyi*, as host plants for *M. oleifera* seedlings, but none of these hosts significantly enhanced the growth of *M. oleifera* seedlings (Li Q. et al., 2019; Xiong et al., 2024). This may be due to the relatively slow growth rate of these hosts, which limits their ability to rapidly promote *M. oleifera* seedling growth. Furthermore, because parasitic plants draw nutrients and water from their host, they can also weaken the growth of the host and can even lead to host death (ArunKumar et al., 2022). Although root hemiparasitic plants are not highly specific in host selection, an ideal host should exhibit fast growth and strong adaptability. Given its rapid growth, tolerance to poor soils, and resilience, *T. diversifolia* appears to meet these criteria, which likely explains the significant growth improvement observed in parasitic *M. oleifera* in this study (S1H treatment).

The host plant used in our experiments, *T. diversifolia*, has a significantly greater effect on the aboveground growth of *M. oleifera* than on its root growth. When comparing the S1N and S1H treatments, the differences in aboveground biomass (fresh and dry weight) of the *M. oleifera* seedlings were extremely significant ($P < 0.001$), while the differences in root biomass were not significant ($P > 0.05$; Figures 1D, E). Moreover, the aboveground biomass of *T. diversifolia* (both fresh and dry weight) was not correlated with the root biomass of *M. oleifera* ($P > 0.05$; Figure 2C). This means that the growth of the above- and belowground parts of *M. oleifera* seedlings in a parasitic state is unbalanced. This uneven growth prevents *M. oleifera* from rapidly developing a large root system, which would enable it to obtain more water and nutrients, as well as from quickly establishing other parasitic connections with other potential host plants, and therefore leads to the seedlings being extremely vulnerable. If the host plant weakens or dies due to parasitism, the *M. oleifera* will likewise weaken or die. In our previous field studies, we found that *M. oleifera* seedlings that initially exhibited good growth often later experienced stagnation or even death, likely due to this unbalanced growth (Personal communication).

4.2 The role of haustoria and roots in supporting the growth of *M. oleifera* under parasitic conditions

Haustoria are specialized structures in parasitic plants that form a physiological bridge with the host, penetrating its tissues and creating a vascular connection to absorb water, nutrients, RNA, proteins, and hormones (Kokla et al., 2022; Shen et al., 2023). The physiology of the haustoria varies significantly before and after the establishment of the parasite-host association (Chen et al., 2021). Biological processes in *M. oleifera* haustoria were significantly enhanced after establishment of the parasite-host association, and of the top 30 GO terms enriched in up-regulated genes, 29 were related to biological processes. (Supplementary Table S3). GO terms enrichment and KEGG pathway analysis of haustoria up-regulated genes (S1N_H vs S1H_H) revealed that the haustoria exhibited enhanced hormone signaling and photosynthesis regulation, responses to pathogens and environmental stresses, secondary metabolism, antioxidant defense, and metabolic pathway regulation. This indicates that the plants are actively adapting to both internal and external growth environments, optimizing their basal metabolism, and strengthening their defense mechanisms and adaptability. Overall, these gene enrichments reflect the plant's ability to coordinate multiple metabolic pathways and signal transduction mechanisms to maintain survival and adaptability under various stresses.

GO term annotations and KEGG pathway analyses indicated that, following the establishment of *M. oleifera*-host associations, the physiology of the roots was also enhanced in response to biotic and abiotic stresses, as well as in photosynthetic activity, energy production, and secondary metabolism. Additionally, the capacities for basic metabolism and material conversion were improved,

including the metabolism of nitrogenous compounds, carbohydrates, and lipids. This indicates that the overall ability of *M. oleifera* to adapt to its environment, utilize nutrients, synthesize essential compounds, and maintain vital functions significantly increased following establishment of the association with the host plant. Such enhancements are crucial for the plant's growth, development, and resilience to environmental stress.

The hormone levels in the haustoria of *M. oleifera* seedlings attached to a host exhibited significant changes, with 13 of the top 30 most enriched GO annotations (S1N-H vs. S1H-H) for up-regulated genes linked to hormone-related physiological activities. Such significant fluctuations in hormone levels within the haustoria upon attaching to the host have been observed in various hemiparasitic plants (Ashapkin et al., 2023). For example, following contact with the host, the expression levels of genes involved in the synthesis and regulation of hormones such as IAA, cytokinins, GAs, ABA, ACC, and JA changed significantly in the haustoria of the root hemiparasite *Santalum album* (Zhang et al., 2015). Our research further confirmed that the levels of these plant hormones in the haustoria of *M. oleifera* exhibited significant increases or decreases following host attachment. Changes in these plant hormones regulate a variety of physiological activities, including the biosynthesis of secondary metabolites, metabolic pathways, plant hormone biosynthesis, hormone signal transduction, zeatin biosynthesis, tryptophan metabolism, circadian entrainment, and more (Figure 5). In root hemiparasitic plants like *M. oleifera*, the regulatory effects of plant hormones play a crucial role in plant development after the haustoria establish a connection with the host. Although our study did not measure the changes in hormone content in *M. oleifera* roots following parasitism, we can infer from the transcriptome data that the hormones in the roots will coordinate with those in the haustoria to regulate plant growth.

During the process of establishing a parasitic relationship between a hemiparasitic plant and its host, the spatiotemporal expression of genes, accumulation of proteins, changes in hormone levels, and the development of the haustorium are all regulated in a coordinated manner. In the process of seed germination and parasitism establishment by *Psittacanthus schiedeanus* on host branches, the expression levels of genes from different categories exhibit a regular and continuous upregulation or downregulation (Ibarra-Laclette et al., 2022). Similarly, during the infection of mesquite by *P. calyculatus*, the activity of cell wall-degrading enzymes shows stage-specific patterns: cellulase and β-1,4-glucosidase play a dominant role during haustorium development, while xylanase, endo-1,4-β-glucanase, and protease are significantly activated during the haustorium penetration and vascular connection stages; Plant hormones, such as auxins and cytokinins, exhibit spatial concentration gradients and are directly involved in the regulation of haustorium development (Aguilar-Venegas et al., 2023). This study primarily monitored the overall growth of the plant and did not specifically observe the stages of haustorium development. Consequently, the current transcriptomic sequencing and hormone content data reflect a composite of the entire haustorium development process, which may dilute the gene expression characteristics and hormone dynamic changes specific to certain developmental stages. However, the gene expression patterns and hormone regulatory networks that consistently promote plant growth should remain relatively stable throughout the monitoring period. The significant upregulation and downregulation of genes in the haustorium and roots of S1H-treated plants (Supplementary Table S7) reflect the overall changes occurring in the root system after parasitism.

4.3 Cultivation of *M. oleifera*

Due to the unique challenges of cultivating parasitic plants, they are rarely grown or introduced through artificial means (Thorogood et al., 2022). However, many parasitic plants have significant and unique economic value, and cultivating these plants for sustainable use remains a challenge in plant cultivation technology (Pignone and Hammer, 2016). It took 150 years of research into *Santalum album*, a representative root hemiparasitic tree, to uncover its host preferences, parasitic growth patterns, haustorial physiological characteristics, and host adaptations to parasitism (ArunKumar et al., 2022). Over more than 40 years of artificial cultivation, the habits of *M. oleifera* have gradually become understood, particularly after the discovery of its semi-parasitic root characteristics. However, despite its significant economic value, sophisticated and effective cultivation practices are still lacking.

To advance the development of *M. oleifera* cultivation techniques, we propose four key measures based on our research, our prior cultivation experience, and other studies on root semi-parasitic plants (1) After planting, *M. oleifera* should be able to quickly establishe a parasitic relationship with the host plant, ensuring that the seedlings acquire the necessary nutrients for early growth. Our study reveals a significant positive correlation between the growth of *M. oleifera* and the number of haustoria attached to the host roots. Other studies have shown that, as nutrient reserves in the roots of *M. oleifera* seedlings gradually deplete over the first growing season (Chen et al., 2024), the seedlings' ability to establish parasitic relationships with other potential host plants also declines. When artificially selecting host plants, it is essential to consider not only the plant species (Matthies, 2017) but also its age (Moncalvillo and Matthies, 2023) and proximity (Thyroff et al., 2023) to *M. oleifera*. (2) The host plant should demonstrate strong environmental adaptability and have minimal growth requirements, ensuring its robust development and creating favorable conditions for the parasitism of *M. oleifera*. This is because the parasitic plant consumes a significant amount of the host plant's water and nutrients. Strong adaptability of the host plant to adverse conditions ensures its normal growth while providing suitable conditions for the parasitic plant (Press et al., 1998; Zhang et al., 2023). In our cultivation of *M. oleifera*, we observed that during drought stress, the host plant wilts and eventually dies, subsequently threatening the survival of *M. oleifera*. (3) The host plant should grow rapidly, be vigorous, and

be relatively larger than the parasitic plant, thereby providing sufficient nutrients to support the parasitism of *M. oleifera*. Our research shows that the aboveground biomass of *M. oleifera* is significantly and positively correlated ($P<0.01$) with that of *T. diversifolia*, and this relationship has been confirmed in other studies as well (Hautier et al., 2010; Matthies, 2017). (4) The host plant should be resilient to top pruning, which can promote the growth of *M. oleifera* by artificially adjusting the distribution of light and nutrients between it and the host plant. The rapid growth of host plants can enhance the growth of parasitic plants; however, excessive growth may block light and deplete nutrients. This contradiction can be managed by artificially pruning the tops of the host plants.

5 Conclusions

The efficiency of *M. oleifera* seedlings in directly absorbing nutrients from the soil is quite low, making it difficult for the plants to grow normally without a host. There is a significant positive correlation between the aboveground biomass of *M. oleifera* seedlings and the biomass of the host plant, indicating that, in order to effectively promote their own growth, *M. oleifera* seedlings must establish a parasitic relationship with a rapidly growing host larger than themselves. Once the seedlings establish a parasitic relationship with the appropriate host, the host plant promotes the growth rate of the aboveground parts of *M. oleifera* seedlings significantly more their belowground parts. This leads to an imbalance in growth between the aboveground and belowground parts of *M. oleifera*, which poses potential a risk to the survival of the hemiparasite if the host weakens or dies.

Transcriptome analysis revealed that parasitism induces significant changes in haustorial physiology, with up-regulated genes predominantly linked to hormone metabolism, stress responses, antibiotic biosynthesis, and lipid metabolism. These changes suggest an active metabolic response to support nutrient acquisition and defense. The hormone levels in haustoria parasitizing other host plant species undergo significant changes, influencing various metabolic processes in *M. oleifera* and ultimately impacting its growth. Moreover, the physiological activity of the roots undergoes significant changes following the establishment of the *M. oleifera*-host association, with a marked enhancement in the hemiparasite's response to biotic and abiotic stresses, as well as improvements in basic metabolic processes. These changes suggest that host attachment strengthened the overall adaptability, nutrient synthesis, and stress resilience in *M. oleifera*, all of which are crucial for its growth, development, and survival. *M. oleifera* seedlings only transform from slow-growing plants into fully developed adults with normal physiological functions after establishing a parasitic relationship with a suitable host.

Data availability statement

The data presented in the study are deposited in the Genome Sequence Archive in National Genomics Data Center, China National Center for Bioinformation / Beijing Institute of Genomics, Chinese Academy of Sciences (accession number GSA: CRA027511) that are publicly accessible at https://ngdc.cncb.ac.cn/gsa.

Author contributions

SW: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Writing – original draft, Writing – review & editing, Visualization. CZ: Data curation, Formal Analysis, Investigation, Writing – original draft. WY: Data curation, Formal Analysis, Investigation, Writing – original draft. JC: Data curation, Formal Analysis, Writing – original draft. MS: Formal Analysis, Investigation, Writing – original draft.

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 National Natural Science Foundation of China (grant number 31860081 and 31560092), the Science Fund of Yunnan Government, China (202001AS070019), and the Science and Technology Basic Resources Investigation Program of China (2017FY100100), Wild Plant Protection Project in Yunnan Province, China.

Acknowledgments

We extend our gratitude to Wen-Zhong Yang for his valuable suggestions on paper revisions, to Zhong-Fu Zhang for his assistance with plant cultivation, to Han-Bing Cai for his support in data analysis, and to Yan-Ling Zheng and Jia-Hui Chen for their guidance in data analysis.

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

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.2025.1589651/full#supplementary-material

SUPPLEMENTARY DATA SHEET 1

Detailed experimental methods for RNA extraction and qRT-PCR.

References

Aguilar-Venegas, M., Quintana-Rodriguez, E., Aguilar-Hernandez, V., López-Garcia, C. M., Conejo-Dávila, E., Brito-Argáez, L., et al. (2023). Protein profiling of *Psittacanthus calyculatus* during mesquite infection. *Plants* 12, 464. doi: 10.3390/plants12030464

ArunKumar, A. N., Joshi, G., Warrier, R. R., and Karaba, N. N. (2022). *Indian Sandalwood: A Compendium* (Singapore: Springer Nature Singapore Pte Ltd.), 229–237.

Ashapkin, V. V., Kutueva, L. I., Aleksandrushkina, N. I., Vanyushin, B. F., Teofanova, D. R., and Zagorchev, L. I. (2023). Genomic and epigenomic mechanisms of the interaction between parasitic and host plants. *Int. J. Mol. Sci.* 24, 2647. doi: 10.3390/ijms24032647

Bell, T. L., and Adams, M. A. (2011). Attack on all fronts: functional relationships between aerial and root parasitic plants and their woody hosts and consequences for ecosystems. *Tree Physiol.* 31, 3–15. doi: 10.1093/treephys/tpq108

Chen, L., Guo, Q., Zhu, Z., Wan, H., Qin, Y., and Zhang, H. (2021). Integrated analyses of the transcriptome and small RNA of the hemiparasitic plant *Monochasma savatieri* before and after establishment of parasite-host association. *BMC Plant Biol.* 21, 90. doi: 10.1186/s12870-021-02861-6

Chen, Q. P., Li, Y. J., Li, Y., Xue, R. J., and Li, A. R. (2024). Dynamics of tissue nutrient content in relation to declining seedling growth in *Malania oleifera*. *Guihaia* 44, 137–146. doi: 10.11931/guihaia.gxzw202303048

Chen, W. D., Wang, P. F., PU, T., Wang, J. W., Pang, Y., and Wang, J. (2022). Symbiotic effect of co-cultivated plants on *Malania oleifera* seedlings. *Acta Agric. Univ. Jiangxiensis.* 44, 1197–1206. doi: 10.13836/j.jjau.2022119

Chen, L., Zhu, Z., Guo, Q., Guo, J., Huang, Z., and Zhang, H. (2020). The changes of morphological and physiological characteristics in hemiparasitic *Monochasma savatieri* before and after attachment to the host pant. *PeerJ* 8, e9780. doi: 10.7717/peerj.9780

Hautier, Y., Hector, A., Vojtech, E., Purves, D., and Turnbull, L. A. (2010). Modelling the growth of parasitic plants. *J. Ecol.* 98, 857–866. doi: 10.1111/j.1365-2745.2010.01657.x

Honaas, L. A., Wafula, E. K., Yang, Z., Der, J. P., Wickett, N. J., Altman, N. S., et al. (2013). Functional genomics of a generalist parasitic plant: Laser microdissection of host-parasite interface reveals host-specific patterns of parasite gene expression. *BMC Plant Biol.* 13, 9. doi: 10.1186/1471-2229-13-9

Huang, K. X., Lai, J. Y., Shi, H. M., Zhuang, J., and Li, X. D. (2008). Research situation on conservation and utilization of Malania oleifera, a single species genus of China. *J. Guangxi. Agric. Biol. Sci.* 27, 76–80.

Ibarra-Laclette, E., Venancio-Rodríguez, C. A., Vásquez-Aguilar, A. A., Alonso-Sánchez, A. G., Pérez-Torres, C. A., Villafán, E., et al. (2022). Transcriptional basis for haustorium formation and host establishment in hemiparasitic *Psittacanthus schiedeanus* mistletoes. *Front. Genet.* 13. doi: 10.3389/fgene.2022.929490

Kokla, A., Leso, M., Zhang, X., Simura, J., Serivichyaswat, P. T., Cui, S., et al. (2022). Nitrogen represses haustoria formation through abscisic acid in the parasitic plant *Phtheirospermum japonicum*. *Nat. Commun.* 13, 2976. doi: 10.1038/s41467-022-30550-x

Lee, S. K. (1980). *Malania*, a new genus of oil-yielding plant. *Bull. Bot. Lab. North-East. For. Inst.* 1, 67–72.

Li, Q., Chen, J., Yu, X., and Gao, J. M. (2019). A mini review of nervonic acid: Source, production, and biological functions. *Food Chem.* 301, 125286. doi: 10.1016/j.foodchem.2019.125286

Li, Y. P., Jing, Y. P., Mao, J. H., Li, Y. P., and Li, S. L. (2019). Root hemiparasitic characteristics of *Malania oleifera*. *J. West. Chin. For. Sci.* 48, 1–12. doi: 10.16473/j.cnki.xblykx1972.2019.04.001

Li, A. R., Li, Y. J., Smith, S. E., Smith, F. A., and Guan, K. Y. (2013). Nutrient requirements differ in two *Pedicularis* species in the absence of a host plant: implication for driving forces in the evolution of host preference of root hemiparasitic plants. *Ann. Bot.* 112, 1099–1106. doi: 10.1093/aob/mct179

Li, A. R., Mao, P., and Li, Y. J. (2019). Root hemiparasitism in *Malania oleifera* (Olacaceae), a neglected aspect in research of the highly valued tree species. *Plant Divers.* 41, 347–351. doi: 10.1016/j.pld.2019.09.003

Li, Y., Zhou, C., Yan, X. J., Zhang, J., and Xu, J. (2016). Simultaneous analysis of ten phytohormones in *Sargassum horneri* by high-performance liquid chromatography with electrospray ionization tandem mass spectrometry. *J. Sep. Sci.* 39, 1804–1813. doi: 10.1002/jssc.201501239

Liu, Z. Q. (1984). Preliminary report on afforestation of *malania oleifera* in rocky mountains. *Yunnan. For. Sci. Technol.* 2, 28–30.

Lv, S. H., Wei, C. Q., Huang, F. Z., He, Y. L., Zeng, D. J., Li, X. K., et al. (2016). Fruit and seed traits and adaptability to rocky desertification mountain of rare tree species Malania oleifera. *Chin. J. Ecol.* 35, 57–62. doi: 10.13292/j.1000-4890.201601.008

Matthies, D. (2017). Interactions between a root hemiparasite and 27 different hosts: Growth, biomass allocation and plant architecture. *Perspect. Plant Ecol. Evol. Syst.* 24, 118–137. doi: 10.1016/j.ppees.2016.12.006

Moncalvillo, B., and Matthies, D. (2023). Host age affects the performance of the root hemiparasitic plant *Rhinanthus alectorolophus*. *Ecol. Evol.* 13, e10167. doi: 10.1002/ece3.10167

Musselman, L. J. (1969). Observations on the life history of *Aureolaria grandiflora* and *Aureolaria pedicularia* (Scrophulariaceae). *Am. Midl. Nat.* 82, 307–311. doi: 10.2307/2423848

Ou, G. Z. (1981). A new presence of important fatty acid (cis-tetracos-15-enoic): Oil of *Malania oleifera* Chun et Lee. *Acta Bot. Yunnan.* 3, 181–184.

Pignone, D., and Hammer, K. (2016). Parasitic angiosperms as cultivated plants? *Genet. Resour. Crop Evol.* 63, 1273–1284. doi: 10.1007/s10722-016-0416-x

Press, M. C., Graves, J. D., and Stewart, G. R. (1998). Transpiration and carbon acquisition in root hemiparasitic angiosperms. *J. Exp. Bot.* 39, 1009–1014. doi: 10.1093/jxb/39.8.1009

Seel, W. E., Parsons, A. N., and Press, M. C. (1993). Do inorganic solutes limit growth of the facultative hemiparasite *Rhinanthus minor* L. in the absence of a host. *New Phytol.* 124, 283–289. doi: 10.1111/j.1469-8137.1993.tb03818.x

Shen, G., Zhang, J., Lei, Y., Xu, Y., and Wu, J. (2023). Between-plant signaling. *Annu. Rev. Plant Biol.* 74, 367–386. doi: 10.1146/annurev-arplant-070122-015430

Speetjens, T. M., and Jacobs, D. F. (2023). Nutrient availability has a greater influence than pot host on seedling development of hemiparasitic Hawaiian sandalwood (*Santalum paniculatum* Hook. and Arn.). *Forests* 14, 1–14. doi: 10.3390/f14030458

Tang, T. F., Liu, X. M., Ling, M., Lai, F., Zhang, L., Zhou, Y. H., et al. (2013). Constituents of the essential oil and fatty acid from *Malania oleifera*. *Ind. Crops Prod.* 43, 1–5. doi: 10.1016/j.indcrop.2012.07.003

Tesitel, J., Plavcová, L., and Cameron, D. D. (2010). Interactions between hemiparasitic plants and their hosts. *Plant Signal. Behav.* 5, 1072–1076. doi: 10.4161/psb.5.9.12563

Thorogood, C., Witono, J. R., Mursidawati, S., and Fleischmann, A. (2022). Parasitic plant cultivation: examples, lessons learned and future directions. *Sibbaldia.: Int. J. Bot. Garden. Hortic.* 21, 109–136. doi: 10.24823/Sibbaldia.2022.1892

Thyroff, E. C., Rose, K. M. E., Idol, T. W., Moon, Q., Burney, O. T., and Jacobs, D. F. (2023). Co-planting of a fast-growing, nitrogen-fixing host tree facilitates regeneration of the root hemiparasitic 'iliahi (Hawaiian sandalwood). *For. Ecol. Manage.* 542, 121084. doi: 10.1016/j.foreco.2023.121084

Wang, S. H., Chen, J., Yang, W., Hua, M., and Ma, Y. P. (2021). Fruiting character variability in wild individuals of *Malania oleifera*, a highly valued endemic species. *Sci. Rep.* 11, 23605. doi: 10.1038/s41598-021-03080-7

Wickett, N. J., Honaas, L. A., Wafula, E. K., Das, M., Huang, K., Wu, B., et al. (2011). Transcriptomes of the parasitic plant family Orobanchaceae reveal surprising conservation of chlorophyll synthesis. *Curr. Biol.* 21, 2098–2104. doi: 10.1016/j.cub.2011.11.011

Xiong, X. D., Liu, A. Z., Lu, S., Guo, X. C., Jia, D. S., and Xu, D. B. (2024). Effects of three host plants on the growth and biomass allocation of *Malania oleifera*. *J. Southwest. For. Univ.* 44, 191–197. doi: 10.11929/j.swfu.202306050

Yang, T., Zhang, R., Tian, X., Yao, G., Shen, Y., Wang, S., et al. (2023). The chromosome-level genome assembly and genes involved in biosynthesis of nervonic acid of *Malania oleifera*. *Sci. Data* 10, 298. doi: 10.1038/s41597-023-02218-8

Zhang, X., Berkowitz, O., Teixeira da Silva, J. A., Zhang, M., Ma, G., Whelan, J., et al. (2015). RNA-Seq analysis identifies key genes associated with haustorial

development in the root hemiparasite *Santalum album*. *F. Plant Sci.* 6. doi: 10.3389/fpls.2015.00661

Zhang, Y. B., Scanlon, M. C., Liu, J. X., Song, X. Y., Yang, D., Zhang, Y. J., et al. (2023). You are what you eat: nutrient and water relations between mistletoes and hosts. *New Phytol.* 238, 567–583. doi: 10.1111/nph.18747

Zhang, X. H., Teixeira da Silva, J. A., Duan, J., Deng, R., Xu, X., and Ma, G. (2012). Endogenous hormone levels and anatomical characters of haustoria in *Santalum album* L. seedlings before and after attachment to the host. *J. Plant Physiol.* 169, 859–866. doi: 10.1016/j.jplph.2012.02.010

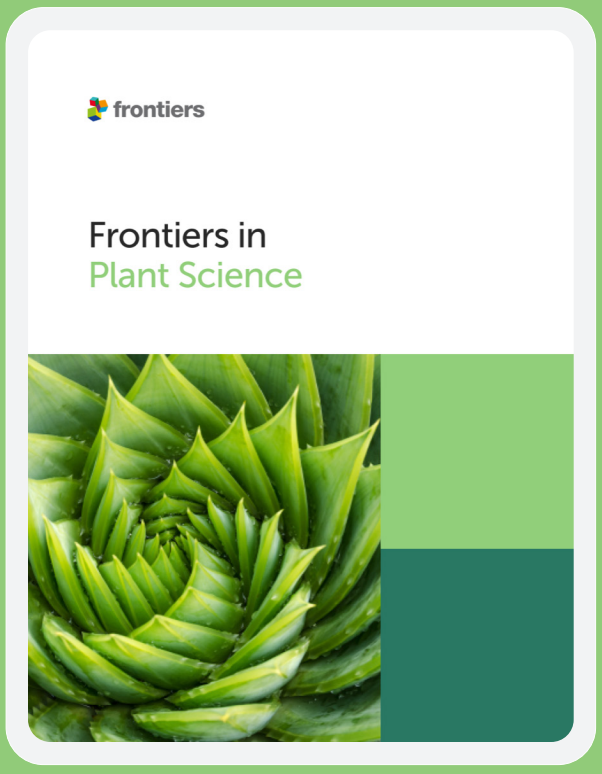
Frontiers in
Plant Science

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