

Genetic advancements for improving the plant tolerance to biotic and abiotic stresses

Edited by

Krishnanand P. Kulkarni, Amaranatha Reddy Vennapusa,
Balaji Aravindhan Pandian and Rupesh Deshmukh

Published in

Frontiers in Genetics



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 978-2-8325-4990-2
DOI 10.3389/978-2-8325-4990-2

About Frontiers

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

Frontiers journal series

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

Dedication to quality

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

What are Frontiers Research Topics?

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

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

Genetic advancements for improving the plant tolerance to biotic and abiotic stresses

Topic editors

Krishnanand P. Kulkarni — Delaware State University, United States
Amaranatha Reddy Vennapusa — Delaware State University, United States
Balaji Aravindhan Pandian — Enko Chem Inc., United States
Rupesh Deshmukh — Central University of Haryana, India

Citation

Kulkarni, K. P., Vennapusa, A. R., Pandian, B. A., Deshmukh, R., eds. (2024). *Genetic advancements for improving the plant tolerance to biotic and abiotic stresses*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-4990-2

Topic editor Dr. Balaji Aravindhan Pandian is employed by Enko Chem Inc. All other Topic Editors declare no competing interests concerning the Research Topic subject.

Table of contents

- 05 **Editorial: Genetic advancements for improving the plant tolerance to biotic and abiotic stresses**
Krishnanand P. Kulkarni, Amaranatha R. Vennapusa, Balaji Aravindhan Pandian and Rupesh Deshmukh
- 09 **Unraveling the genomic reorganization of polygalacturonase-inhibiting proteins in chickpea**
Vishnutej Ellur, Wei Wei, Rishikesh Ghogare, Shyam Solanki, George Vandemark, Robert Brueggeman and Weidong Chen
- 24 **Characterization and gene expression patterns analysis implies BSK family genes respond to salinity stress in cotton**
Yuqian Lei, Yupeng Cui, Ruifeng Cui, Xiugui Chen, Junjuan Wang, Xuke Lu, Delong Wang, Shuai Wang, Lixue Guo, Yuexin Zhang, Cun Rui, Yapeng Fan, Mingge Han, Lanjie Zhao, Hong Zhang, Xiaoyu Liu, Nan Xu, Jing Wang, Hui Huang, Xixian Feng, Yanlong Xi, Kesong Ni, Menghao Zhang, Tiantian Jiang and Wuwei Ye
- 35 **Comparative transcriptomic analysis provides key genetic resources in clove basil (*Ocimum gratissimum*) under cadmium stress**
Bin Wang, Yukun Wang, Xiao Yuan, Yuanyuan Jiang, Yunna Zhu, Xinmiao Kang, Jinming He and Yanhui Xiao
- 51 ***De novo* sequencing, assembly, and characterization of *Asparagus racemosus* transcriptome and analysis of expression profile of genes involved in the flavonoid biosynthesis pathway**
Chanchal Malik, Sudhanshu Dwivedi, Tilahun Rabuma, Ravinder Kumar, Nitesh Singh, Anil Kumar, Rajesh Yogi and Vinod Chhokar
- 66 **Discovery of genomic regions associated with grain yield and agronomic traits in Bi-parental populations of maize (*Zea mays*. L) Under optimum and low nitrogen conditions**
Collins Kimutai, Noel Ndlovu, Vijay Chaikam, Berhanu Tadesse Ertiro, Biswanath Das, Yoseph Beyene, Oliver Kiplagat, Charles Spillane, Boddupalli M. Prasanna and Manje Gowda
- 84 **Genome-wide identification of long non-coding RNAs and their potential functions in radish response to salt stress**
Xiaochuan Sun, Mingjia Tang, Liang Xu, Xiaobo Luo, Yutong Shang, Weike Duan, Zhinan Huang, Cong Jin and Guodong Chen
- 97 **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**
Dennis O. Omondi, Mathews M. Dida, Dave K. Berger, Yoseph Beyene, David L. Nsibo, Collins Juma, Suresh L. Mahabaleswara and Manje Gowda

- 113 **Advancements and prospects of CRISPR/Cas9 technologies for abiotic and biotic stresses in sugar beet**
Varucha Misra, A. K. Mall, Himanshu Pandey, Santeshwari Srivastava and Avinash Sharma
- 140 **The role of WRKY transcription factors in exogenous potassium (K⁺) response to NaCl stress in *Tamarix ramosissima***
Yahui Chen, Xuanyi Zhang, Yunlong Fan, Dezong Sui, Jiang Jiang and Lei Wang
- 151 **QTL mapping for seedling and adult plant resistance to stripe and leaf rust in two winter wheat populations**
Alma Kokhmetova, Nagenahalli Dharmegowda Rathan, Deepmala Sehgal, Angelina Malysheva, Madina Kumarbayeva, Makpal Nurzhuma, Ardak Bolatbekova, Gopalareddy Krishnappa, Elena Gultyaeva, Asia Kokhmetova, Zhenis Keishilov and Kanat Bakhytuluy
- 168 **Unraveling the regulatory network of miRNA expression in Potato Y virus-infected of *Nicotiana benthamiana* using integrated small RNA and transcriptome sequencing**
Hongping Song, Xinwen Gao, Liyun Song, Yubing Jiao, Lili Shen, Jinguang Yang, Changquan Li, Jun Shang, Hui Wang, Songbai Zhang and Ying Li
- 181 **A comprehensive review on *Gossypium hirsutum* resistance against cotton leaf curl virus**
Sahar Nadeem, Syed Riaz Ahmed, Tahira Luqman, Daniel K. Y. Tan, Zahra Maryum, Khalid Pervaiz Akhtar, Sana Muhy Ud Din Khan, Muhammad Sayyam Tariq, Nazar Muhammad, Muhammad Kashif Riaz Khan and Yongming Liu



OPEN ACCESS

EDITED AND REVIEWED BY
Andrew H. Paterson,
University of Georgia, United States

*CORRESPONDENCE
Krishnanand P. Kulkarni,
✉ kkulkarni@desu.edu

[†]These authors have contributed equally to this work

RECEIVED 01 May 2024
ACCEPTED 13 May 2024
PUBLISHED 27 May 2024

CITATION

Kulkarni KP, Vennapusa AR, Pandian BA and Deshmukh R (2024), Editorial: Genetic advancements for improving the plant tolerance to biotic and abiotic stresses. *Front. Genet.* 15:1426680. doi: 10.3389/fgene.2024.1426680

COPYRIGHT

© 2024 Kulkarni, Vennapusa, Pandian and Deshmukh. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). 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: Genetic advancements for improving the plant tolerance to biotic and abiotic stresses

Krishnanand P. Kulkarni^{1*†}, Amaranatha R. Vennapusa^{1†},
Balaji Aravindhan Pandian² and Rupesh Deshmukh³

¹Department of Agriculture and Natural Resources, Delaware State University, Dover, DE, United States, ²Enko, Mystic, CT, United States, ³Department of Biotechnology, Central University of Haryana, Mahendragarh, India

KEYWORDS

genetic advancement, abiotic stress, disease resistance, quantitative trait loci, genome-wide association studies, molecular mechanisms

Editorial on the Research Topic

Genetic advancements for improving the plant tolerance to biotic and abiotic stresses

Introduction

Climate change is multi-faceted and primarily includes rising temperatures, increasing frequency of extreme weather events, enhanced accumulation of greenhouse gases (e.g., CO₂, methane) in the atmosphere, and alterations in precipitation patterns (Gray and Brady, 2016; Vennapusa et al., 2023). Such events exaggerate abiotic stress factors while providing favorable situations to biotic stressors such as pests and diseases. Therefore, understanding and safeguarding the complex interplay of these stressors is imperative for developing resilient crop varieties and ensuring global food and nutritional security (Kulkarni et al., 2018). In this regard, plant scientists face a major challenge to develop strategies for enhancing crop resilience and ensuring food security.

In nature, plants are concurrently exposed to multiple abiotic stress factors (Nabi et al., 2019), which allow them to co-evolve and develop endurance through a wide range of finely balanced responses (Lima et al., 2015; Gonzalez Guzman et al., 2022). Understanding the molecular, genetic, and regulatory mechanisms involved in plant's responses will help devise strategies to mitigate climate change. Advancements in next-generation sequencing techniques led to the development of high-quality reference genomes, high-throughput genotyping systems, and complex genetic linkage maps, which enabled the precise identification of genomic regions associated with a trait of interest through genome-wide association studies (GWAS) and quantitative trait loci (QTL) mapping (Asekova et al., 2021; Uffelmann et al., 2021). Researchers are able to significantly accelerate the pace of crop genetic improvements for simple and complex traits through marker-assisted selection (MAS) or genomic selection (GS). Owing to these developments, substantial progress has been made in understanding the mechanisms of plant tolerance to abiotic and biotic stresses and adaptation. With recent advances in gene-editing technologies, it is now feasible to develop plants with

favorable traits, such as increased productivity, improved adaptability to abiotic and biotic stresses, and better nutritional composition (Wang et al., 2017; Movahedi et al., 2023).

In this Research Topic on “Genetic advancements for improving the plant tolerance to biotic and abiotic stresses”, we have collated 12 articles (10 original research articles and two review articles) that highlight the potential insights into the physiological, molecular, and genetic factors involved in imparting abiotic and biotic stress tolerance in plants. This compilation highlights many recent findings, methods, and molecular genetic resources for crop improvement programs and contributes to enhancing agricultural productivity and sustainability.

Understanding plant resistance to biotic stresses

Plants must constantly defend themselves against attacks from various kinds of pathogens, causing destructive diseases. Five articles from this Research Topic describe the effective genetic and molecular mechanisms plants have developed to recognize and respond to infection by several pathogens. The study by Omondi et al. identified genomic regions associated with gray leaf spot (GLS) and northern corn leaf blight (NCLB) resistance in tropical maize (*Zea mays* L.). The QTLs explained 64.2% and 64.9% of the total phenotypic variance for GLS and NCLB resistance, respectively. Further, association mapping identified 11 significant single nucleotide polymorphisms (SNPs) associated with GLS and 16 SNPs with NCLB resistance; many of these SNPs found to be co-localized with QTL regions identified in a biparental population. Using the same sets of plants, the authors tested genomic prediction models and highlighted the application of GS for improving resistance to multiple diseases in maize.

For genetic dissection of plant resistance to yellow rust (YR) and leaf rust (LR), Kokhmetova et al. evaluated two recombinant inbred line populations for multiple years and identified nine and four stable QTLs for adult plant resistance to YR and LR, respectively. Further *in silico* analysis of these regions revealed candidate genes that regulate host response toward pathogen infection. The stable QTLs and genes identified in this study are helpful genetic resources for developing rust-resistant varieties through MAS.

Cotton (*Gossypium* sp.) production is continuously being challenged by various biotic and abiotic stresses, among which cotton leaf curl disease (CLCuD) is most prevalent and causes severe losses. The review by Nadeem et al. comprehensively addresses the challenges faced by CLCuD in cotton crop cultivation, strategies for disease management, and potential options to improve the resistance. The review highlights the basis of CLCuV strains and genetic architecture, transmission vectors and mechanisms, and the potential of advanced biotechnological and molecular breeding approaches, such as genetic engineering, speed breeding, and genome editing, in developing CLCuD-resistant cotton varieties. Overall, the review emphasizes adopting these advancements to revolutionize crop improvement and ensure sustainable cotton production.

The pioneering study by Ellur et al. provides the first comprehensive characterization of specific chickpea (*Cicer*

arietinum) Polygalacturonase-inhibiting proteins (PGIPs) using computational, localization, and gene expression studies. Identifying CaPGIP3 and CaPGIP4 on chromosome 3 in this study expands the genomic understanding of CaPGIPs, dictating a reconsideration of their genomic organization from previous reports on chromosome 6. The findings highlight the potential of *CaPGIP1*, *CaPGIP3*, and *CaPGIP4* genes in combating chickpea pathogens, specifying their structural and functional similarities with other legume PGIPs.

Song et al. investigated the role of microRNAs (miRNAs) in combating Potato virus Y (PVY) disease, a major threat to crop quality and productivity. The comparative study of small RNA and transcriptome sequencing on healthy and PVY-infected *Nicotiana benthamiana* tissues revealed the differentially expressed miRNAs and predicted their potential gene targets. This inclusive analysis enhances our understanding of a complex regulatory network involved in the host's response to PVY infection and connection with miRNAs, offering valuable insights into novel strategies for PVY control.

Understanding plant tolerance to abiotic stresses

Recent advancements in understanding plants' adaptive responses to abiotic stress have opened new avenues for improving stress tolerance and productivity under adverse environmental conditions. The articles related to abiotic stress in the current Research Topic provide an overview of key developments driving progress in crucial areas of plant science. The review by Misra et al. highlights the potential use of CRISPR/Cas 9 technology for abiotic and biotic stress management in sugar beet (*Beta vulgaris*). Candidate - means potential candidate genes, known for their role in alkaline, cold, and heavy metal stresses, can be precisely modified via CRISPR/Cas 9 technology to enhance sugar beet's resilience to abiotic stresses with minimal off-target effects. The authors present a list of potential genes that can be targeted by a CRISPR tool for genetic improvements for heat, drought, heavy metal, and other abiotic stresses, as well as biotic stress factors such as cyst nematode (*Heterodera schachtii schmidt*) resistance, necrotic yellow vein virus resistance, and insect-pest resistance.

Salt stress affects many plant life stages, from seed germination to flowering and fruiting. Sun et al. studied salt stress influencing yield and quality in radish (*Raphanus sativus* L.). They specifically targeted the role of long non-coding RNAs (lncRNAs) in regulating salt stress response using the genome-wide identification study. Notably, this is the first comprehensive analysis of salt stress-responsive lncRNAs in radish and provides potential insights into the regulatory networks governing salt stress tolerance and their potential target sites in radish. These findings lay a foundation for understanding stress responses and future functional characterization of lncRNAs associated with radish defense response against high salinity. Transcription factors also play pivotal regulatory and molecular switch roles by controlling the expression of downstream genes or directly acting to protect plants from the damage induced by salt stress. Chen et al. analyzed the role of WRKY genes in *Tamarix ramosissima* in response to NaCl stress.

Many WRKY genes actively upregulated their expression levels to resist NaCl stress through the plant-pathogen interaction and the MAPK signaling pathways. The authors reported that WRKY transcription factors can modulate the activation or inhibition of related genes during NaCl stress, which may enhance plant's adaptability to saline environments and mitigate NaCl-induced damage.

Brassinosteroid is a phytohormone reported to play an essential role in acclimation to environmental stresses, resistance to pathogens, and cell elongation, resulting in increased crop yield and plant growth (Kim and Russinova, 2020). Lei et al. characterized brassinosteroid signaling kinase (BSK) family genes that respond to salinity stress in cotton to decipher the functions of BSK genes on growth and responses to abiotic stresses and learn the evolutionary relationship of cotton BSKs. The study showed the induced expression of BSK genes under salt stress, suggesting the important role of BSK gene regulation in cotton growth, development, and stress response.

Inadequate soil fertility (particularly low nitrogen) is a primary cause of low maize yields in sub-Saharan Africa, leading to food insecurity, malnutrition, and rural poverty in smallholder farming communities. Kimutai et al. discovered QTLs associated with grain yield and agronomic traits in maize (*Zea mays* L.) under optimum and low nitrogen conditions to address this issue. These QTLs are of significant value for further validation and possible rapid introgression into maize populations using MAS. The authors further implied that using GS can improve genetic gain in maize breeding for low nitrogen stress tolerance.

The research by Malik et al. elucidates the genes involved in flavonoid biosynthesis pathways through *de novo* transcriptome sequencing of *A. racemosus*, commonly called Shatavari. The authors identified several differentially expressed genes, including the NAC family of plant transcription factors known for their role in abiotic stress tolerance in plants. The study provides a molecular understanding of the genes involved in the flavonoid biosynthesis pathway for further secondary metabolic improvement in the *Asparagus racemosus*.

The study by Wang et al. investigated the impact of cadmium (Cd) stress on clove basil (*Ocimum gratissimum* L.) physiological traits and integrated it with transcriptome analysis to explore the molecular-level responses. The findings provide novel candidate genes and genetic resources for understanding Cd stress adaptation mechanisms and offer insights into potential strategies for crop breeding programs to improve heavy metal (HM) tolerance. This research highlights the importance of integrating the physiognomic studies with genomics in aromatic plants like clove basil, which play a significant role in safeguarding HM pollution while enhancing aromatic essential oil production and thus advancing phytoremediation efforts and sustainable utilization of HM-contaminated soils.

References

- Asekova, S., Oh, E., Kulkarni, K. P., Siddique, M. I., Lee, M. H., Kim, J. I., et al. (2021). An integrated approach of QTL mapping and genome-wide association analysis identifies candidate genes for phytophthora blight resistance in sesame (*Sesamum indicum* L.). *Front. Plant Sci.* 12, 604709. doi:10.3389/fpls.2021.604709
- Gonzalez Guzman, M., Cellini, F., Fotopoulos, V., Balestrini, R., and Arbona, V. (2022). New approaches to improve crop tolerance to biotic and abiotic stresses. *Physiol. Plant.* 174 (1), e13547. doi:10.1111/ppl.13547
- Gray, S. B., and Brady, S. M. (2016). Plant developmental responses to climate change. *Dev. Biol.* 419 (1), 64–77. doi:10.1016/j.ydbio.2016.07.023
- Kim, E.-J., and Russinova, E. (2020). Brassinosteroid signalling. *Curr. Biol.* 30 (7), R294–R298. doi:10.1016/j.cub.2020.02.011
- Kulkarni, K. P., Tayade, R., Asekova, S., Song, J. T., Shannon, J. G., and Lee, J.-D. (2018). Harnessing the potential of forage legumes, alfalfa, soybean, and cowpea for sustainable agriculture and global food security. *Front. Plant Sci.* 9, 1314. doi:10.3389/fpls.2018.01314

Conclusion

The contributions to this Research Topic provide a catalog of physiological, molecular, and genetic information related to plants' tolerance to abiotic and biotic stresses. The identified stress adaptive mechanisms, potential QTLs, candidate genes, and state-of-the-art approaches mentioned in these articles broaden our knowledge and provide genetic resources and avenues for crop improvement programs toward enhanced productivity and climate resilience. This Research Topic addressed vital factors and advancements that will accelerate genetic improvements in producing climate-resilient plants.

Author contributions

KK: Writing–original draft, Writing–review and editing. AV: Writing–original draft, Writing–review and editing. BP: Writing–review and editing. RD: Writing–review and editing.

Acknowledgments

We thank all the authors for their excellent contributions and reviewers for their constructive comments and input, which made this Research Topic successful.

Conflict of interest

Author BP is employed by Enko, Mystic, CT, United States.

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.

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.

- Lima, J. M., Nath, M., Dokku, P., Raman, K., Kulkarni, K., Vishwakarma, C., et al. (2015). Physiological, anatomical and transcriptional alterations in a rice mutant leading to enhanced water stress tolerance. *AoB Plants* 7, plv023. doi:10.1093/aobpla/plv023
- Movahedi, A., Aghaei-Dargiri, S., Li, H., Zhuge, Q., and Sun, W. (2023). CRISPR variants for gene editing in plants: biosafety risks and future directions. *Int. J. Mol. Sci.* 24 (22), 16241. doi:10.3390/ijms242216241
- Nabi, R. B. S., Tayade, R., Hussain, A., Kulkarni, K. P., Imran, Q. M., Mun, B.-G., et al. (2019). Nitric oxide regulates plant responses to drought, salinity, and heavy metal stress. *Environ. Exp. Bot.* 161, 120–133. doi:10.1016/j.envexpbot.2019.02.003
- Uffelmann, E., Huang, Q. Q., Munung, N. S., De Vries, J., Okada, Y., Martin, A. R., et al. (2021). Genome-wide association studies. *Nat. Rev. Methods Prim.* 1 (1), 59. doi:10.1038/s43586-021-00056-9
- Vennapusa, A. R., Nimmakayala, P., Zaman-Allah, M. A., and Ratnakumar, P. (2023). Editorial: physiological, molecular and genetic perspectives of environmental stress response in plants. *Front. Plant Sci.* 14, 1213762. doi:10.3389/fpls.2023.1213762
- Wang, M., Mao, Y., Lu, Y., Tao, X., and Zhu, J.-k. (2017). Multiplex gene editing in rice using the CRISPR-Cpf1 system. *Mol. Plant* 10 (7), 1011–1013. doi:10.1016/j.molp.2017.03.001



OPEN ACCESS

EDITED BY

Balaji Aravindhan Pandian,
Enko Chem Inc., United States

REVIEWED BY

Roy Kirsch,
Max Planck Institute for Chemical
Ecology, Germany
Silvio Tundo,
University of Padua, Italy
Karthik Putta,
ENKOCHEM, United States

*CORRESPONDENCE

Weidong Chen,
✉ weidong.chen@usda.gov

RECEIVED 19 March 2023

ACCEPTED 26 May 2023

PUBLISHED 05 June 2023

CITATION

Ellur V, Wei W, Ghogare R, Solanki S,
Vandemark G, Brueggeman R and
Chen W (2023), Unraveling the genomic
reorganization of polygalacturonase-
inhibiting proteins in chickpea.
Front. Genet. 14:1189329.
doi: 10.3389/fgene.2023.1189329

COPYRIGHT

© 2023 Ellur, Wei, Ghogare, Solanki,
Vandemark, Brueggeman and Chen. This
is an open-access article distributed
under the terms of the [Creative
Commons Attribution License \(CC BY\)](#).
The use, distribution or reproduction in
other forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does not
comply with these terms.

Unraveling the genomic reorganization of polygalacturonase-inhibiting proteins in chickpea

Vishnutej Ellur¹, Wei Wei², Rishikesh Ghogare³, Shyam Solanki⁴,
George Vandemark⁵, Robert Brueggeman⁶ and Weidong Chen^{5*}

¹Molecular Plant Science, Washington State University, Pullman, WA, United States, ²Department of Plant Pathology, Washington State University, Pullman, WA, United States, ³Department of Horticultural Sciences, Texas A&M University, College Station, TX, United States, ⁴Department of Agronomy, Horticulture and Plant Science, South Dakota State University, Brookings, SD, United States, ⁵Grain Legume Genetics Physiology Research, Pullman, WA, United States, ⁶Department of Crop and Soil Science, Washington State University, Pullman, WA, United States

Polygalacturonase-inhibiting proteins (PGIPs) are cell wall proteins that inhibit pathogen polygalacturonases (PGs). PGIPs, like other defense-related proteins, contain extracellular leucine-rich repeats (eLRRs), which are required for pathogen PG recognition. The importance of these PGIPs in plant defense has been well documented. This study focuses on chickpea (*Cicer arietinum*) PGIPs (CaPGIPs) owing to the limited information available on this important crop. This study identified two novel CaPGIPs (CaPGIP3 and CaPGIP4) and computationally characterized all four CaPGIPs in the gene family, including the previously reported CaPGIP1 and CaPGIP2. The findings suggest that CaPGIP1, CaPGIP3, and CaPGIP4 proteins possess N-terminal signal peptides, ten LRRs, theoretical molecular mass, and isoelectric points comparable to other legume PGIPs. Phylogenetic analysis and multiple sequence alignment revealed that the CaPGIP1, CaPGIP3, and CaPGIP4 amino acid sequences are similar to the other PGIPs reported in legumes. In addition, several cis-acting elements that are typical of pathogen response, tissue-specific activity, hormone response, and abiotic stress-related are present in the promoters of *CaPGIP1*, *CaPGIP3*, and *CaPGIP4* genes. Localization experiments showed that CaPGIP1, CaPGIP3, and CaPGIP4 are located in the cell wall or membrane. Transcript levels of *CaPGIP1*, *CaPGIP3*, and *CaPGIP4* genes analyzed at untreated conditions show varied expression patterns analogous to other defense-related gene families. Interestingly, CaPGIP2 lacked a signal peptide, more than half of the LRRs, and other characteristics of a typical PGIP and subcellular localization indicated it is not located in the cell wall or membrane. The study's findings demonstrate CaPGIP1, CaPGIP3, and CaPGIP4's similarity to other legume PGIPs and suggest they might possess the potential to combat chickpea pathogens.

KEYWORDS

polygalacturonase inhibitory proteins (PGIPs), gene family, defense-related gene, biotic stress response, leucine-rich repeats (LRRs), promoter analysis, constitutive gene expression, subcellular localization

1 Introduction

Plants deploy a variety of barriers to withstand numerous pathogenic stresses, one of which is the cell wall, a physical barrier that serves as the first line of defense. Pathogens produce enzymes known as cell wall-degrading enzymes (CWDEs) to overcome this plant barrier (Kubicek et al., 2014). Pectin-degrading enzymes called polygalacturonases are among the most important CWDEs. The middle lamella is the plant cell's outermost layer that connects the primary cell walls of adjacent cells (Daher and Braybrook, 2015). Middle lamella is rich in pectin, which primarily constitutes homogalacturonan, a linear homopolymer of D-galacturonic acid monomers linked by α -(1–4) glycosidic linkage (Mohnen, 1999). Pectin determines the integrity and rigidity of plant tissue (Voragen et al., 2009), and degrading pectin enables quick access to the components within the cell. By breaking down glycosidic linkages between D-galacturonic acid residues, PGs degrade homogalacturonan and subsequently pectin causing cell separation and maceration of the host tissue (Kalunke et al., 2015; Mojsov, 2016). Polygalacturonases (PG) are secreted at the early stages of infection (De Lorenzo et al., 2001). In defense, plants use polygalacturonase inhibiting proteins (PGIPs) to impede PGs' pectin-depolymerizing activity. Plant PGIPs are located on the cell wall and their potential to suppress PG activity is correlated with plant disease resistance (Ge et al., 2019).

PGIPs are highly conserved proteins (Di Matteo et al., 2003). So far, PGIPs have been reported in every characterized plant species or mutant (Kalunke et al., 2015). Most PGIPs are generally intronless, except a few that include a short intron (Kalunke et al., 2015). PGIPs, like many other resistance gene products, contain extracellular type leucine-rich repeats (eLRRs) (Di Matteo et al., 2003; Kalunke et al., 2014). PGIPs are composed of 10 incomplete LRRs of approximately 24 residues each, which are arranged into two β -sheets. β 1 occupies the inner concave side of the molecules, while β 2 occupies the outer convex side. These repeats form β -sheet/ β -turn/ α -helix containing LRR motifs. Motifs that occupy the β 1 inner concave side are critical for interaction with PGs.

Albersheim and Anderson, 1971 were the first to report PGIP gene activity in 1971. The first PGIP gene, however, was isolated in French beans 20 years later (Toubart et al., 1992). Several PGIP genes have been identified in several crops based on sequence identity since 1971. Except for some members belonging to Brassicaceae (Hegedus et al., 2008), most PGIP genes do not undergo large expansion and may exist as single gene per genome (Di Giovanni et al., 2008), or clustered into small gene families (Ferrari et al., 2003). In legumes, PGIP genes have been characterized in *Glycine max*, *Medicago sativa*, *Medicago truncatula*, *Phaseolus acutifolius*, *Phaseolus coccineus*, *Phaseolus lunatus*, *Phaseolus vulgaris*, *Pisum sativum*, and *Vigna radiata* (Veronico et al., 2011; Gao and Gui, 2015; Kalunke et al., 2015; Matsaunyan et al., 2015; Kaewwongwal et al., 2017). However, only *M. sativa*, *M. truncatula*, *V. radiata*, *P. vulgaris*, and *G. max*'s genome have more than one PGIP gene (Kalunke et al., 2015; Wang et al., 2022).

Pathogen PG inhibition by PGIPs is well established. PGIPs like other defense molecules can be used against pathogens and pests (Hamera et al., 2014; Zhou et al., 2017). The majority of the identified legume PGIPs inhibited fungal infections, such as *G. max*'s GmPGIP7 (D'Ovidio et al., 2006; Frati et al., 2006;

Kalunke et al., 2014), *M. truncatula*'s MtPGIP1, MtPGIP2 (Song and Nam, 2005), *P. vulgaris*'s PvPGIP1, PvPGIP 2, PvPGIP3, PvPGIP 4 (Desiderio et al., 1997; D'Ovidio et al., 2006; Frati et al., 2006), *P. acutifolius*'s PaPGIP2, *P. coccineus*'s PcPGIP2, *P. lunatus*'s PIPGIP2 (Farina et al., 2009), and *Brassica napus*, BnPGIPs (Wang et al., 2021). However, PGIPs from *V. radiata*, VrPGIP1, and VrPGIP2 (Kaewwongwal et al., 2017; Zhang et al., 2021) and *Brassica rapa* ssp. *pekinensis* BrPGIPs (Haeger et al., 2021) have shown to inhibit insects and *P. sativum* PsPGIP inhibited nematodes (Veronico et al., 2011).

Some PGIP genes are expressed in untreated conditions when plants are not stressed, while others respond to external cues. Pathogens and pests such as fungi, oomycetes, insects, and nematodes are known to induce PGIP gene expression, as are phytohormones such as abscisic acid (ABA), indole-3-acetic acid (IAA), salicylic acid (SA), and jasmonic acid (JA) (Ferrari et al., 2003; Hwang et al., 2010; Hou et al., 2015). PGIP gene expression is also triggered by wounding and oligogalacturonic acid treatments (Ferrari et al., 2003; Di Matteo et al., 2006). PGIP genes/gene families expression is tissue-specific and developmentally regulated (Li and Smigocki, 2016), studies conducted with basal transcript levels of *B. napus* PGIPs (Hegedus, et al., 2008), *P. vulgaris* PGIPs (Kalunke et al., 2011), and *C. papaya* PGIPs (Broetto et al., 2015) indicate PGIPs are expressed in untreated conditions when plants are not stressed.

Currently eighteen PGIPs have been either computationally or biochemically characterized from nine legume species, but major legumes such as chickpeas, peanuts, and lentil PGIPs remain uninvestigated. This study focuses on PGIPs of chickpeas because chickpeas are the world's second most widely produced and consumed leguminous crop, chickpeas have a high protein content (up to 40% protein by weight), are an excellent source of essential vitamins such as riboflavin, niacin, thiamin, folate, and the vitamin A precursor β -carotene, and have other potential health benefits such as lowering cardiovascular, diabetic, and cancer risks (Jukanti et al., 2012; Sharma et al., 2013; Merga and Haji, 2019). Previous publications only report that the chickpea genome harbors two PGIPs (CaPGIP1 and CaPGIP2) on chromosome 6 (Kalunke et al., 2014; Kalunke et al., 2015). Therefore the goal of this study is to investigate PGIPs in chickpeas to gain a better understanding of their structural features, functional domains, regulatory elements, and genomic organization. CaPGIP genes were cloned, and their sequence features were evaluated in this study. The basal expression of all CaPGIPs was explored. Our findings revealed that CaPGIPs, like other legume PGIPs, had similar characteristics and can play an essential role in plant resistance against pathogens and pests.

2 Materials and methods

2.1 Sequence acquisition, phylogeny, and bioinformatics analysis

To identify PGIP homologs in the chickpea genome, a homology search was performed against the NCBI database using the amino acid sequences of previously known legume PGIPs. SignalP 5.0 was used to identify the presence of signal peptides in the candidate genes identified by the NCBI homology search (Armenteros et al., 2019). The molecular

weight and isoelectric point (pI) were determined using ExPASy Server (Gasteiger et al., 2005). NetNGlyc version 1.0 server was used to analyze the putative N-linked glycosylation sites (Gupta & Brunak, 2002). The Swiss-Model server was used to build homology-based 3D models of *CaPGIPs* (Waterhouse et al., 2018). Protein sequences were aligned using ClustalW through the MEGA X program (Kumar et al., 2018). Jalview was used for multiple sequence alignment with a conservation index of 50% (Waterhouse et al., 2009). A phylogenetic tree was generated using MEGA X (Kumar et al., 2018) with the neighbor-joining phylogenetic statistical method, Poisson model and other settings retained at default. The tree was bootstrapped 1000 times for robustness and *Cucumis sativus* PGIPs (CsPGIPs) were used as outgroup. MEGA X generated trees were visualized using the iTOL version 6.1.1 online tool (Letunic and Bork, 2021). The 1,500 bp upstream sequence for all *CaPGIP* sequences was analyzed for the putative cis-acting regulatory DNA elements using New PLACE (Higo et al., 1999).

2.2 Plant materials

Chickpea (*Cicer arietinum*) cultivar Dwelley was grown in greenhouse conditions. Plants maintained in the greenhouse at $22^{\circ}\text{C} \pm 2^{\circ}\text{C}$. Leaf, stem, root, flower, pod, and seed tissues were collected at different chickpea growth stages, which are mentioned in Table 1. Tissue samples (100 mg) were collected in three biological replicates and were immediately snap-frozen in liquid nitrogen and stored at -80°C .

2.3 Cloning and sequencing

The total RNA was extracted from the leaves of chickpea cultivar Dwelley using the RNeasy Plant Mini Kit (Qiagen, Hilden,

Germany) from 100 mg samples in accordance with the manufacturer's protocol. First strand cDNA synthesis and genomic DNA elimination were performed simultaneously using 5X All-In-One RT MasterMix, containing AccurT Genomic DNA Removal (Applied Biological Materials Inc., Richmond, Canada). cDNA samples were stored at -80°C until use. Full-length ORFs were amplified with Phusion[®] High-Fidelity DNA Polymerase (NEB, Ipswich, MA, United States) using gene-specific primer pairs (Supplementary Table S1) using the following protocol: initial denaturation at 98°C for 30 s and 35 cycles of 98°C for 10 s, 60°C and 72°C for 30 s each and followed by a final elongation at 72°C for 8 min. Amplified PCR products with appropriate expected sizes were purified with the Monarch[®] DNA Gel Extraction Kit (NEB). Purified PCR products were cloned into the pMiniT 2.0 vector (NEB) and transformed into DH10B high-efficiency *E. coli* competent cells (NEB). The plasmids were recovered from *E. coli* using PureYield[™] Plasmid Miniprep (Promega, Madison, WI, United States), verified using Sanger sequencing (Laboratory of Biotechnology & Bioanalysis, Pullman, WA, United States), and compared to the GenBank sequences of *CaPGIP1* (XM_004504675), *CaPGIP3* (XM_004493500), and *CaPGIP4* (XM_012713804).

2.4 RNA isolation, cDNA conversion and quantitative real-time-PCR

To determine the absolute expression of *CaPGIP* genes in untreated conditions, total RNA was extracted from various chickpea tissues from the cultivar Dwelley at different growth stages (Table 1). Based on the timing of infection by the major chickpea fungal pathogens, four growth stages (V1, V6, R1, and R4) were selected (Mazur et al., 2002; West et al., 2003; Bretag and

TABLE 1 Chickpea tissues collected during different chickpea growth stages for absolute gene expression quantification.

Growth stage	Stage	Stage description	Tissue collected	Days after sowing
Vegetative growth stage	V1	First multifoliate leaf has unfolded from the stem	Leaf	10
			Root	10
			Stem	10
	V6	Sixth multifoliate leaf has unfolded from the stem	Leaf	25
			Root	25
			Stem	25
Reproductive growth stage	R1	Early bloom, one open flower on the plant	Flower	55
			Leaf	55
			Root	55
			Stem	55
	R4	Flat pod, pod has reached its full size and is largely flat	Leaf	80
			Pod	80
			Root	80
			Stem	80
			Seed	80

Horsham, 2004; Markel et al., 2008; Jiménez-Fernández et al., 2011; Moore et al., 2011; Khan et al., 2012; Wunsch, 2014; Knights and Hobson, 2016). RNA was extracted from 100 mg samples using the RNeasy Plant Mini Kit (Qiagen) according to the manufacturer's protocol. First-strand cDNA synthesis and genomic DNA elimination were performed simultaneously using 5X All-In-One RT MasterMix, containing AccurT Genomic DNA Removal (Applied Biological Materials, Inc.). To preserve sample integrity, RNA extraction, genomic DNA removal, and cDNA synthesis were performed on the same day. Samples were stored at -80°C until use. Using Primer3Plus (Untergasser et al., 2007) at the default parameters, quantitative PCR primers (Supplementary Table S1) were generated based on the sequences of *CaPGIP1*, *CaPGIP3*, and *CaPGIP4* and the reference gene 18SrRNA and 25SrRNA. The NCBI Primer-BLAST (ncbi.nlm.nih.gov/tools/primer-blast/) program was used to ensure that primers are unique specifically to the corresponding gene. Primers were referenced against the chickpea genome ASM33114v1. Standard curves generated by serial dilution of cDNA for 18SrRNA, 25SrRNA, *CaPGIP1*, *CaPGIP3*, and *CaPGIP4* were used to evaluate primer efficiency (Supplementary Figure S1). Transcript levels of chickpea *PGIPs* (*CaPGIP1*, *CaPGIP3*, and *CaPGIP4*) in chickpea at untreated conditions were evaluated following the Minimum Information for Publication of Quantitative Real-Time PCR Experiments (MIQE) guidelines (Bustin et al., 2009). Each RT-qPCR reaction consisted of 1x SsoAdvancedTM universal Inhibitor-Tolerant SYBR[®] Green Supermix (Bio-Rad, Hercules, CA, United States), 2.5 μM of each gene-specific primer, and cDNA converted from 100 ng RNA in a final reaction volume of 10 μL . No template control (NTC), no amplification control (NAC), and negative reverse transcription (NRT) controls were included for each primer pair, and all reactions were performed with three separate biological replicates in technical triplicates. qPCR was carried out in the CFX96TM Real-Time PCR Detection System using a two-step amplification and melt curve method with the following protocol: 95°C for 3 min, followed by 40 cycles of 95°C for 10 s; 60°C for 30 s; and 72°C for 30 s. Melt curve readings were taken from 65.0°C to 95.0°C with an increment of 0.5°C every 5 s. The absolute gene expression assays were performed by constructing standard curves of the corresponding cloned coding region of *CaPGIPs* (Wong and Medrano, 2005). The expression Ct values of *CaPGIPs* were normalized against the expression Ct values of reference genes 18SrRNA and 25SrRNA. Quantification was done using the relative standard curve method (Supplementary Figure S1) (Pfaffl, 2001). The *CaPGIP* expression values are given as the mean of the normalized expression values of *CaPGIPs* normalized against reference genes 18SrRNA and 25SrRNA. Obtained *CaPGIP* data is shown as gene copy number/microgram of RNA (Forlenza et al., 2012). The statistical significance was determined by one-way analysis of variance (ANOVA) followed by Tukey's *ad hoc* testing using the PROC GLM program in Statistical Analysis System (SAS) version 9.4 (SAS Institute Inc, 2013).

2.5 Subcellular localization

DeepLoc-1.0 (Almagro Armenteros et al., 2017) was used to predict subcellular localization based on *CaPGIP* protein

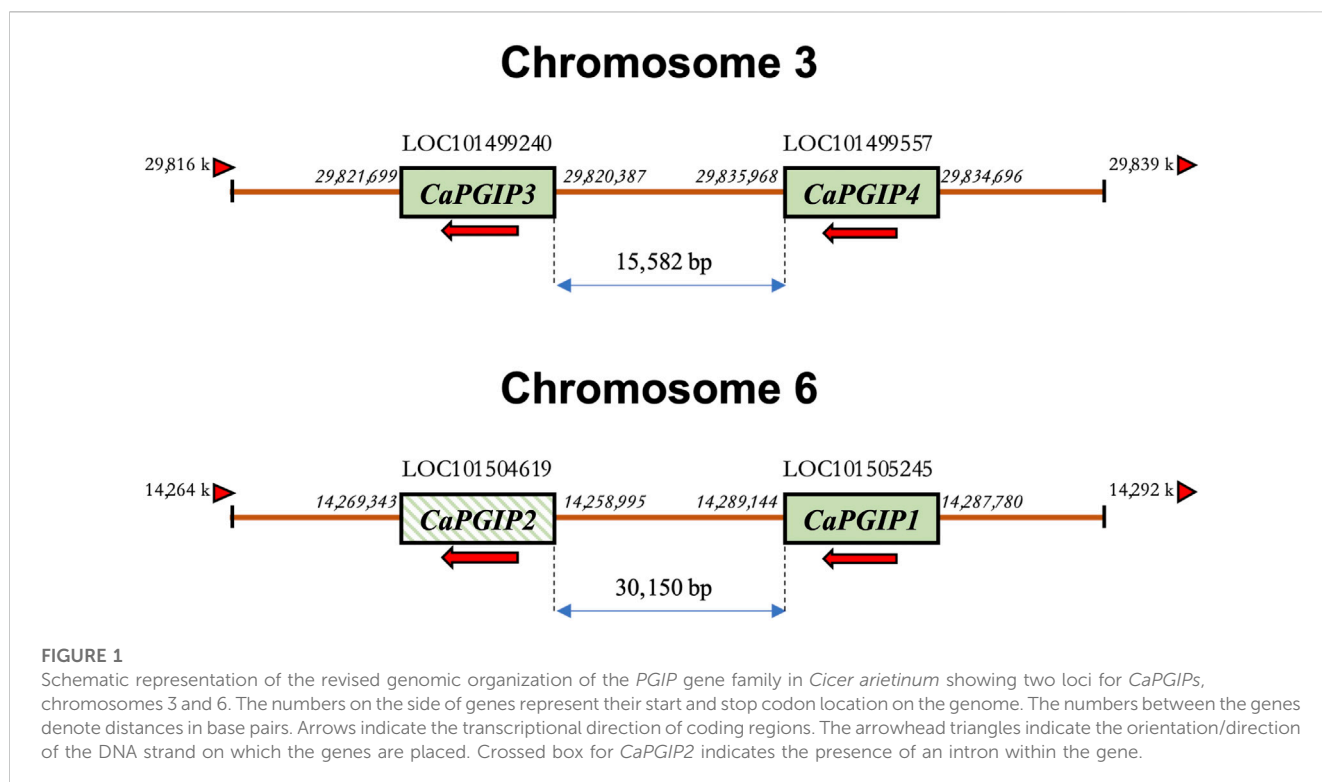
sequences. The complete coding sequences (CDS) of *CaPGIP1*, *CaPGIP2*, *CaPGIP3*, and *CaPGIP4* genes were cloned into pEarleyGate 103 (ABRC) using the gateway cloning approach to determine their subcellular localization. After sequencing validation, these gateway plant expression vectors were transformed into *Agrobacterium tumefaciens* strain EHA105. Using a blunt syringe, transformed EHA105 cultures harboring *CaPGIP*-mGFP plasmids were infiltrated into 4-week-old *N. benthamiana* leaves, two biological replicates along with two technical replicates per gene were infiltrated and imaged. A laser scanning confocal microscope (Leica SP-8) was used to examine and capture the fluorescence emitted by fusion proteins 72 h after infiltration. GFP fluorescence was excited at 488 nm.

3 Results

3.1 Insilco characterization of *CaPGIPs*

CaPGIP1 (LOC101505245) and *CaPGIP2* (LOC101504619) are two previously reported chickpea *PGIPs* (Kalunke et al., 2014; Kalunke et al., 2015). They occupy a 30,150 bp region on chromosome 6. *CaPGIP1* has a single exon with no intron as seen in several legume *PGIPs*. While *CaPGIP2* has two exons separated by a 9,825 bp intron. A homology search against the NCBI chickpea genome assembly ASM33114v1 using the amino acid sequences of known legume *PGIPs* revealed the presence of additional five candidate *PGIP* sequences (Supplementary Table S2). Only two of them, LOC101499240 and LOC101499557, were suitable for designation as prospective *PGIPs* since they were the appropriate size, had a signal peptide, and had ten LRR sequences. They were named *CaPGIP3* and *CaPGIP4* respectively. *CaPGIP3* and *CaPGIP4* are composed of a single exon with no introns and span a 15,582 bp region on chromosome 3. In addition to the previously known locus of *PGIP* genes (Kalunke et al., 2014; Kalunke et al., 2015), our analysis identifies a new locus. The occurrence of two *PGIP* loci in chickpeas, chromosome 3 and chromosome 6, necessitates a new genomic organization of chickpea *PGIP* genes (Figure 1). Full-length cDNA size (bp), ORF size (bp), predicted protein size (aa), predicted signal peptide size (aa), theoretical molecular mass (kDa), and pI for all four *CaPGIPs* are presented in Table 2.

Sequence analysis was conducted for all four *CaPGIPs*. Predicted proteins of *CaPGIP1*, *CaPGIP3*, and *CaPGIP4* exhibited a typical *PGIP* sequence identity. SignalP 5.0 (Armenteros et al., 2017) projected 36, 22, and 20 amino acid signal peptides for *CaPGIP1*, *CaPGIP3*, and *CaPGIP4*, respectively (Table 2). As illustrated in Figure 2, the predicted mature protein sequences for *CaPGIP1*, *CaPGIP3*, and *CaPGIP4* featured an N-terminal domain, a central LRR domain, and a C-terminal domain. These tandemly repeated LRRs fold into a characteristic curved and elongated *PGIP* shape. As observed in homology 3D models generated using PvPGIP2 as a template for *CaPGIP1*, *CaPGIP3*, and *CaPGIP4* (Figure 3), the secondary and tertiary structures of *CaPGIP1*, *CaPGIP3*, and *CaPGIP4* indicate that all 10 LRRs contain an β -turn motif (xxLxLxx) that folds into β 1 sheets. β 1 sheets occupy the *PGIP* scaffold's inner concave face, which is the site for *PG*



interaction. Aside from that, all LRRs have $\beta 2$ sheets and 3_{10} -helices. PGIPs are glycoproteins with N-glycosylation sites ((N-x-S/T; where N is asparagine, x can be any amino acid except proline (P), S is serine, and T is threonine). As a result, the NetNGlyc version 1.0 server predicted five, three, and eight N-glycosylation sites in CaPGIP1, CaPGIP3, and CaPGIP4 proteins, respectively. CaPGIPs contain several conserved cysteine residues. CaPGIP1 has eight, CaPGIP3 has ten, and CaPGIP4 has nine cysteine residues.

Interestingly, CaPGIP2 (Figures 2, 3), on the contrary, lacked many of the above sequence identities. The absence of signal peptide suggested it is not a secretory protein. Its secondary structure reveals that it lacked more than half of the LRR modules, with just the 6th to 10th LRRs. This short LRR sequence on CaPGIP2's C-terminal perfectly matches CaPGIP1's C-terminal. The CaPGIP2 homology 3D model indicated the absence of the distinctive concave face that harbors PG interaction sites (Figure 3). CaPGIP2 has one N-glycosylation site and five cysteine residues, fewer than its counterparts.

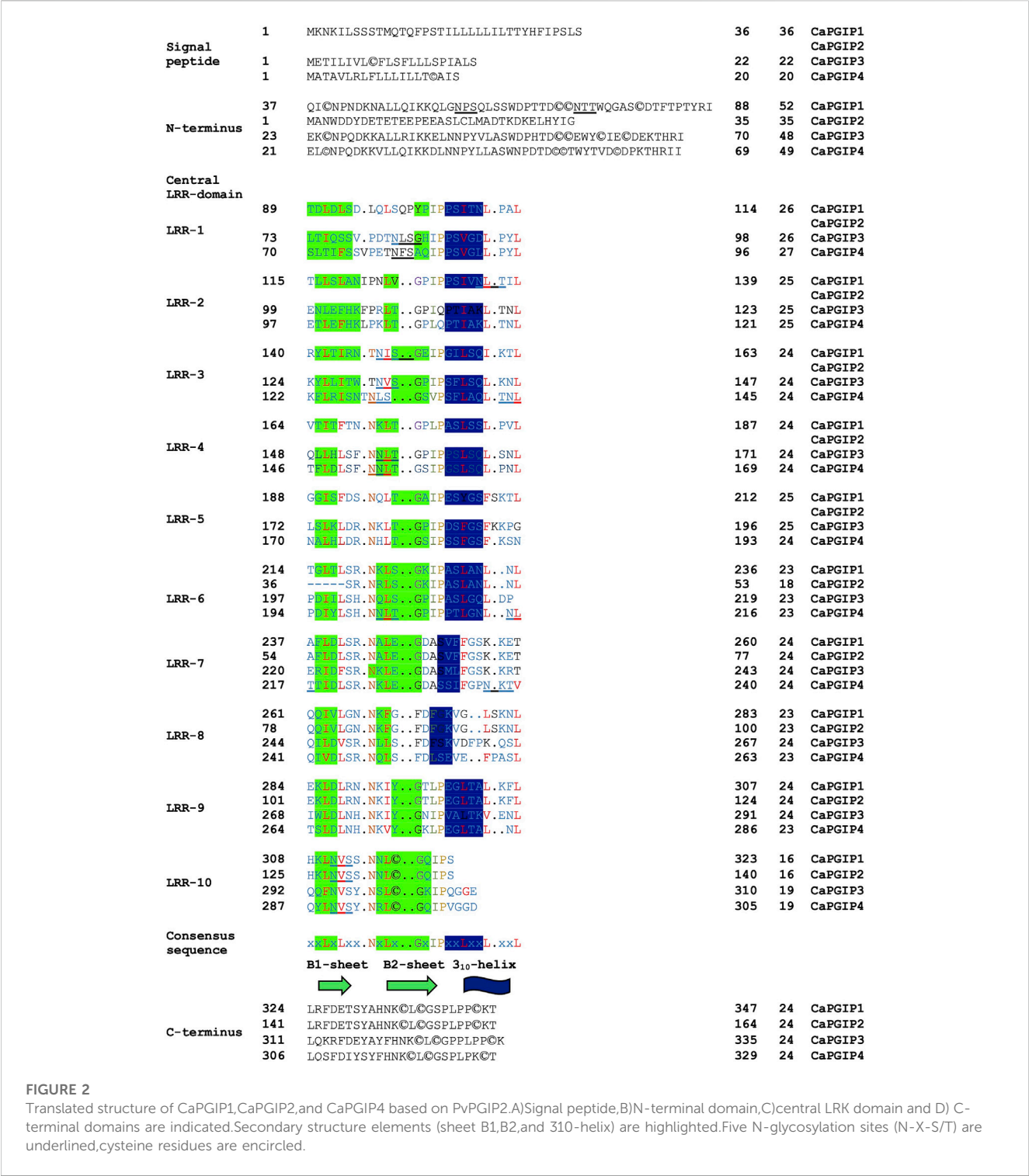
3.2 Sequence comparison and phylogenetic analysis of the CaPGIP proteins

Multiple sequence alignment showed that CaPGIP amino acid sequences are highly similar to those of other legumes such as soybean, common bean, runner bean, tepary bean, lima bean, barrel clover, peas, mung bean, and alfalfa, with the presence of five conserved cysteine residues shared by all. The higher similarity was observed in the $\beta 2$ -sheet regions, along with variable portions present in both β -sheets, as evident in plant-specific LRR proteins (Figure 4). This sequence alignment demonstrates that PGIPs are highly conserved within the legume family.

In the phylogenetic analysis (Figure 5), CaPGIPs were compared to 45 other known PGIPs from various crop families. The tree (Figure 5) comprises five main branches that are separated into monocots and dicots. The Poaceae family is represented by one cluster, the majority of the legume PGIPs are represented by a second cluster, three PGIPs from *Beta vulgaris* are represented by a third cluster. CaPGIP3, and PGIPs from *V. radiata* are represented

TABLE 2 NCBI accession number, full-length cDNA size (bp), ORF size (bp), predicted protein size (aa), predicted signal peptide size (aa), theoretical molecular mass (kDa), and pI for all four CaPGIPs.

Name	NCBI accession	Gene symbol	cDNA (bp)	ORF (bp)	Protein (aa)	Signal peptide (aa)	Molecular mass (kDa)	Isoelectric point (pI)
CaPGIP1	XM_004504675	LOC101505245	1365	1041	347	36	37.65	9.05
CaPGIP2	XM_027335237	LOC101504619	523	492	164	NA	18.18	5.64
CaPGIP3	XM_004493500	LOC101499240	1313	1005	335	22	37.61	8.44
CaPGIP4	XM_012713804	LOC101499557	1273	987	329	20	36.19	6.64



by a fourth cluster, and the remaining PGIPs from other families, such as Actinidiaceae, Apocynaceae, Brassicaceae, Caricaceae, Cucurbitaceae, and Malvaceae, form the final cluster. CaPGIP1 and CaPGIP2 are members of the Leguminosae cluster and exhibit significant similarities to pea PGIP, PsPGIP1. CaPGIP3 and CaPGIP4 are outside the Leguminosae cluster, with CaPGIP3 sharing a high degree of similarity with its other legume counterparts, *V. radiata* PGIPs, VrPGIP1, and VrPGIP2. CaPGIP4 is in a separate cluster that includes PGIPs from various plant families.

3.3 Promoter analysis of *CaPGIPs*

To locate regulatory DNA elements, the 1,500 bp upstream sequence for all *CaPGIP* genes was analyzed. The TATA box and

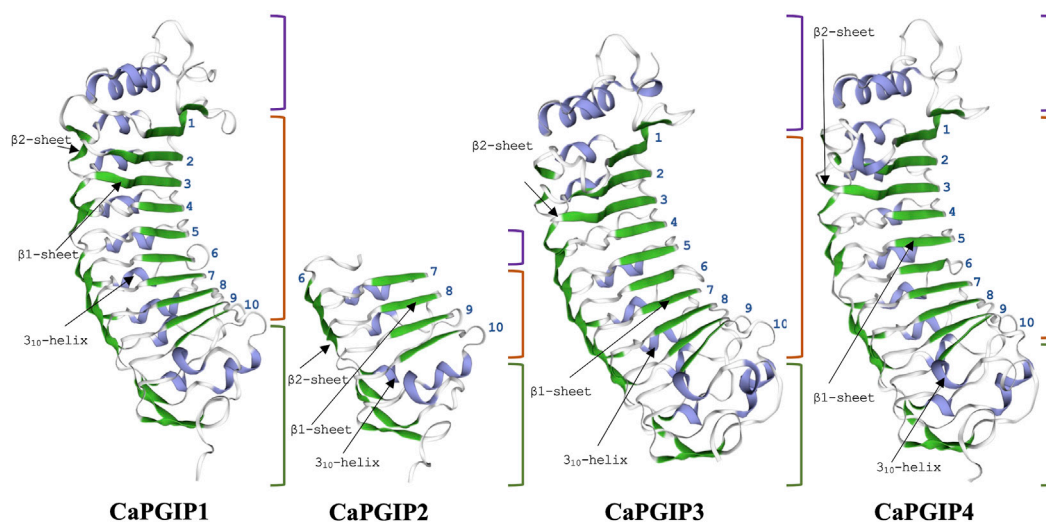


FIGURE 3
Homology 3D model of CaPGIP1, CaPGIP2, CaPGIP3 and CaPGIP4 using PvPGIP2 as a template, b1, and b2 sheets are indicated by green color. 310-helices are indicated with purple color. purple-colored brackets indicate N-terminal domains, orange-colored brackets indicate central LRR domain with LRR numbers marked and green-colored brackets indicate C-terminal domain.

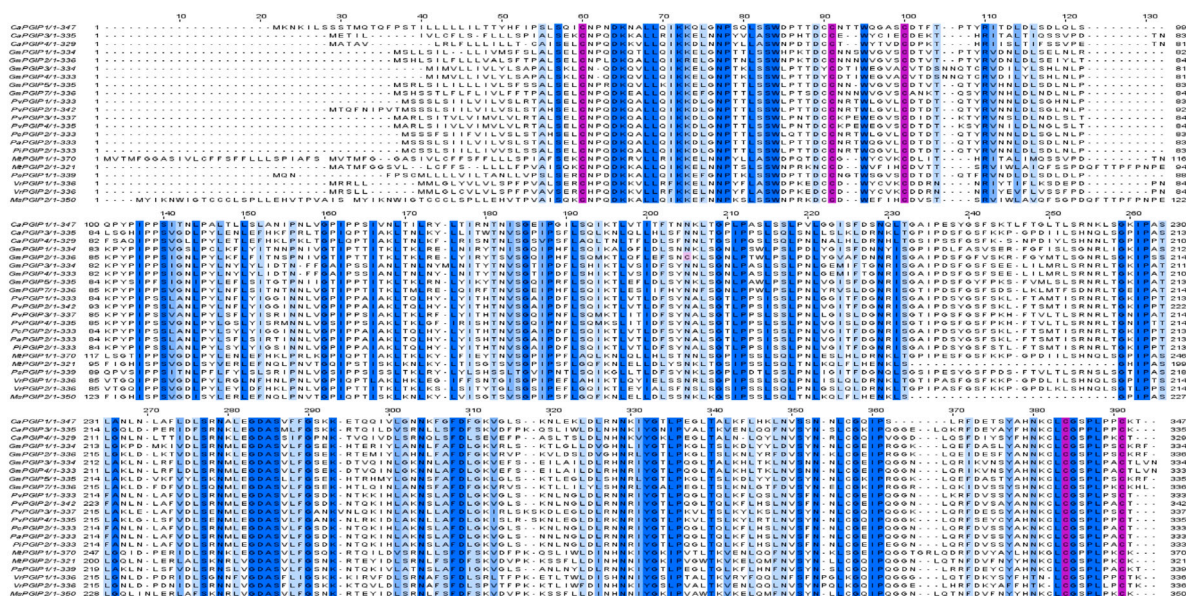
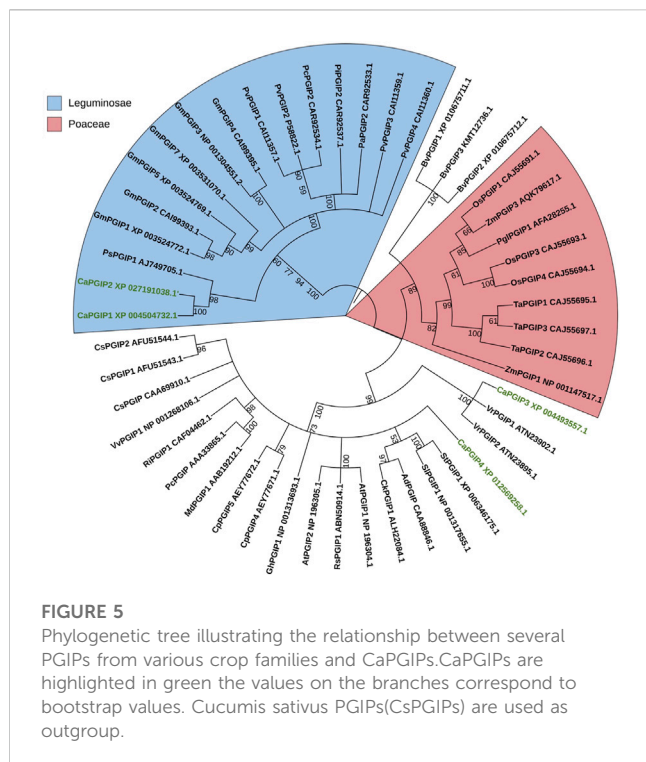


FIGURE 4
Multiple sequence alignment of chickpea PGIPs [CaPGIP1(XP_004504732.1), CaPGIP3(XP_004493557.1), CaPGIP4(XP_012569258.1)], with other legume PGIPs such as soybean [GmPGIP1(XP_003522477.1), GmPGIP2(CAI99393.1), GmPGIP3 (NP_001304551.2), GmPGIP4(CAI99395.1), GmPGIP5(XP_003524769.1), GmPGIP7(XP_003531070.1)], common bean [PvPGIP1(CAI11357.1), PvPGIP2(P58822.10), PvPGIP3(CAI11359.1), PvPGIP4(CAI11360.1)], runner bean [PcPGIP2(CAR92534.1)], tepary bean [PAPGIP2(CAR92533.1)], lima bean [PiPGIP2(CAR92537.1)], barrel clover [MtPGIP1(XP_003625218.1), MtPGIP2(XP_024626259.1)], peas [PsPGIP1(AJ749705.1)], mung bean [VrPGIP1(ATN23902.1), VrPGIP2(ATN23895.1)], and alfalfa [MsPGIP2(ALX18673.1)]. Blue color bar indicates conservation more than 50%, brighter the bar more the conservation, pink color bar indicates cysteine residues.

CAAT box motifs were discovered close to the start codons. CAAT and TATA box sequences were found at -40 and -229 upstream of *CaPGIP1*, respectively, and at -23 and -55 upstream of *CaPGIP3*.

The TATA box was located at a -30 position in the *CaPGIP4* upstream sequence, whereas CAAT was at a -37 position. Elements/motifs associated with plant responses to hormones such as abscisic



acid, gibberellic acid, jasmonic acid, and salicylic acid were also identified. Motifs for wounding response were identified as well. Crucially, numerous elements associated with pathogenicity responses were identified in the promoter regions of the *CaPGIPs*. Table 3 lists these putative cis-acting regulatory elements, as well as their locations and roles. Aside from the motifs mentioned above, additional cis-elements known to mediate tissue-specific activity and plant physiological processes were identified. Stress-related cis-acting regulatory elements associated with drought, dehydration, water, high light, and low-temperature stress, were also identified. All these elements are listed in Supplementary Table S3.

3.4 Cloning and characterization of *CaPGIPs*

CaPGIP genes were cloned and sequenced from the chickpea cultivar “Dwelley”. *CaPGIP1* sequence matched GenBank sequence XM 004504675. However, a C was replaced by a T at the 720th nucleotide position, which was a synonymous substitution with no change in the coded amino acid. *CaPGIP3* sequence matched the GenBank sequence XM 004493500, and *CaPGIP4* sequence matched the GenBank sequence XM 012713804. Transcripts for *CaPGIP2* could not be amplified even with different sets of primers; hence, all subsequent investigations focused on *CaPGIP1*, *CaPGIP3*, and *CaPGIP4*.

3.5 Subcellular localization of *CaPGIPs*

DeepLoc-1.0 uses sequencing information to predict the subcellular localization of plant proteins. Based on the presence

of signal peptides, it was inferred that *CaPGIP1*, *CaPGIP3*, and *CaPGIP4* were secretory and classified as extracellular proteins. *CaPGIP2*, on the other hand, was predicted to be found in the mitochondrion and cytoplasm. To validate those predictions, *Agrobacterium* cells carrying binary vectors of *CaPGIPs* and GFP fusions were infiltrated into *N. benthamiana* leaves for transient expression of encoded proteins in leaf mesophyll and epidermal cells. According to the excitation curves, the fluorescence of *CaPGIPs*-GFP fusion proteins was the same as GFP. *CaPGIP1*, *CaPGIP3*, and *CaPGIP4* fluorescence was visible on the cell boundaries (Figure 6). As a result, they were most likely found on the cell wall or plasma membrane. *CaPGIP2*-GFP fluorescence was seen inside the cell in the cytoplasm and endoplasmic reticulum, and *CaPGIP2* was most likely found in the cytoplasm and endoplasmic reticulum (Figure 6).

3.6 Absolute quantification of *CaPGIP*'s transcripts

CaPGIP transcript levels were investigated at four growth stages using the indeterminate type of Kabuli chickpea variety Dwelley, which matures in 110–120 days. RT-qPCR was utilized to determine the absolute *CaPGIPs* expression levels (Table 1). Transcripts for *CaPGIP1*, *CaPGIP3*, and *CaPGIP4* were ubiquitously detected in all the studied tissues. In the vegetative stages V1 and V6 stages were investigated (Figure 7). *CaPGIP1*, *CaPGIP3*, and *CaPGIP4* transcript levels were higher in the V1 leaf compared to the root and stem. In V1 leaf *CaPGIP4* expressed significantly higher, followed by *CaPGIP3*, then *CaPGIP1*. For stem in the V1 stage, similar to leaf *CaPGIP4* expressed significantly higher, followed by *CaPGIP3*, then *CaPGIP1*. Roots in the V1 stage had a different expression pattern, where *CaPGIP3* expressed significantly higher, followed by *CaPGIP1*, then *CaPGIP4*. *CaPGIPs* expression was lower in the V6 vegetative stage compared to V1. In contrast to V1, the transcript levels of *CaPGIP1*, *CaPGIP3*, and *CaPGIP4* were higher in the V6 stem compared to the leaf and root. In V6 stem *CaPGIP4* expressed significantly higher, followed by *CaPGIP3*, then *CaPGIP1*. For roots in the V6 stage, *CaPGIP4* and *CaPGIP3* expression levels were statistically similar and higher than *CaPGIP1*. For V6 leaves, *CaPGIP1* and *CaPGIP3* expression was statistically similar and lower than *CaPGIP4*.

R1 and R4 were investigated in the reproductive stages, and the transcript levels were analyzed for flower, pod, and seed along with leaf, stem, and root (Figure 8). All *CaPGIPs* in leaves showed higher expression levels in the R1 stage compared to other tissues. In the R1 leaf, *CaPGIP3* expressed significantly higher, followed by *CaPGIP4*, then *CaPGIP1*. *CaPGIP1* and *CaPGIP4* expression levels were statistically similar and lower than *CaPGIP3* in the R1 root. For stem and flowers in R1, *CaPGIP4* had a higher expression, followed by *CaPGIP3* and then *CaPGIP1*. Similar to the R1 stage, all *CaPGIPs* in R4 leaves expressed at higher levels than other tissues. In the R4 leaf, *CaPGIP4* expressed significantly higher, followed by *CaPGIP3*, then *CaPGIP1*. In R4 roots, *CaPGIP3* expressed significantly higher, followed by *CaPGIP1*, then *CaPGIP4*. For the R4 stem, *CaPGIP1* and *CaPGIP3* expression levels were statistically similar and lower than *CaPGIP4*. For

TABLE 3 Putative hormonal and pathogenesis related *cis*-acting regulatory elements identified in the promoter regions of *CaPGIPs*.

<i>Cis</i> - element	Position			Signal sequence	Function	References
	<i>CaPGIP1</i>	<i>CaPGIP3</i>	<i>CaPGIP4</i>			
ABRELATERD1	—	287(–),288(+),1415(–)	561(–),1132(–),1167(+)	ACGTG	Abscisic acid response	Nakashima et al., 2006
ABRERATCAL	—	286(–),287(+),1414(–)	560(–),1131(–),1166(+)	MACGYGB	Abscisic acid response	Kaplan et al., 2006
ACGTABRE	—	—	559(–)	ACGTGKC	Abscisic acid response	Hattori et al., 2002
ARFAT	—	—	857(+)	TGTCTC	Abscisic acid response	Nag et al., 2005
ARR1AT	320(–), 600(+),622(+), 637(–),969(–),990(–), 1008(–),1023(–),1054(–), 1223(–)	151(–),1293(–),1410(–), 1435(+),1488(–)	98(–),113(+),190(–),245(–), 355(+),371(–),872(+),896(–), 978(+),1106(+),1193(–), 1196(+),1275(–), 1360(+), 1446(–)	NGATT	Bacterial response	Ross et al., 2004
BIHD1OS	198(–),1337(–)	483(+),661(+),809(–), 1044(–)	1300(–),1369(+)	TGTCA	Pathogen response	Luo et al., 2005
GADOWNAT	—	—	559(–)	ACGTGTC	Abscisic acid response	Nakashima et al., 2006
GAREAT	123(+),144(+)	—	166(+),741(–)	TAACAAR	Gibberellic acid response	Ogawa et al., 2003
GT1CONSENSUS	209(+),236(+),283(+),356(+),402(–),422(–), 430(–),714(+),754(–), 843(+),1091(–),1092(–), 1130(–),1146(–), 1203(+),1369(–)	26(+),33(+),54(–), 73(+),87(+),391(–), 429(+),477(+),575(–), 612(+),668(–),712(+), 1102(–),1116(–),1136(–), 1310(–),1332(+),1465(–)	53(+),66(–),86(–),106(–), 279(+),314(+),330(–),787(–), 799(–),907(+),908(+), 1021(–),1022(–),1031(–), 1046(–),1456(–),1470(–)	GRWAAW	Salicylic acid response	Buchel et al., 1999
GT1GMSCAM4	209(+),422(–),1091(–), 1146(–),1203(+),1369(–)	26 (+),87(+),391(–), 612(+),1136(–)	53(+),314(+)	GAAAAA	Pathogen response	Park et al., 2004
		1310(–),1332(+)	799(–),1470(–)			
SEBFCONSSTPR10A	—	660(+),783(–)	856(+)	YTGTCWC	Pathogen response	Boyle & Brisson, 2001
T/GBOXATPIN2	—	1415(–)	561(–),1132(–),1166(+)	AACGTG	Jasmonate response	Boter et al., 2004
WBOXNTERF3	—	—	412(–), 780 (–),1060(–), 1119 (+), 1293 (+), 1307 (+)	TGACY	Wound response	Nishiuchi et al., 2004

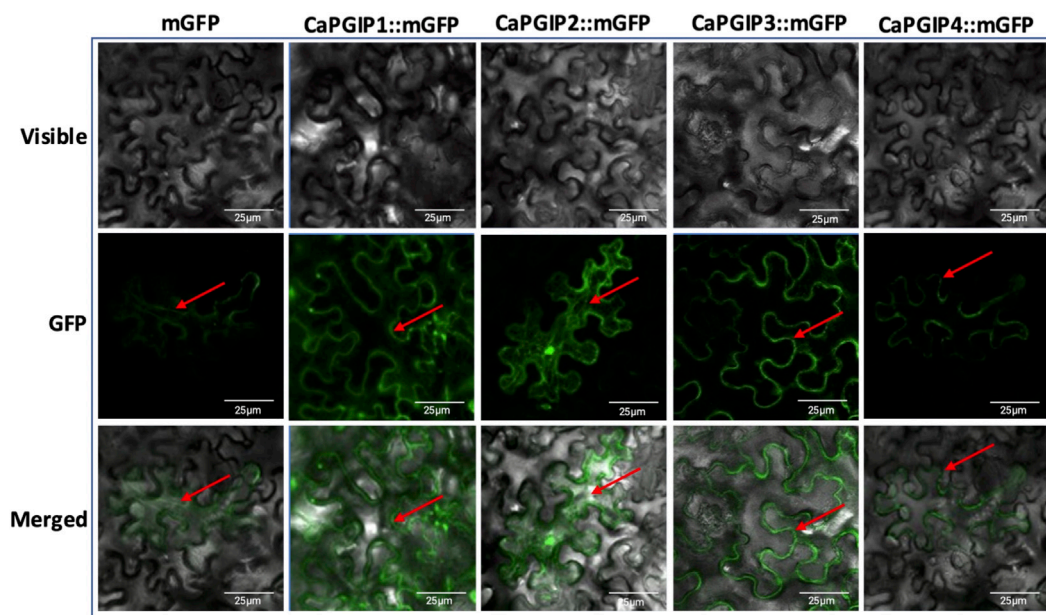


FIGURE 6

Subcellular localization of Capgips. Full-length Capgips fused with a green fluorescent protein (GFP) are transiently expressed in *Nicotiana benthamiana* leaves by agroinfiltration. The images show the fluorescence emitted by fusion proteins was captured 72 h after infiltration using a laser scanning confocal microscope as mGFP fluorescence in green color, visible light in brightfield images, merged as merged images of mGFP and visible. Cells transformed with mGFP is the control. Two biological replicates along with two technical replicates per gene were infiltrated and imaged. Red arrows indicate localization.

Pods and seeds in R4, *CaPGIP4* had a higher expression, followed by *CaPGIP3* and then *CaPGIP1*. Overall, the highest expression levels for all *CaPGIPs* were found in the leaves. Additionally, *CaPGIP4* showed the highest expression among the three genes, with levels about one and a half times higher than *CaPGIP3* and double the expression of *CaPGIP1*.

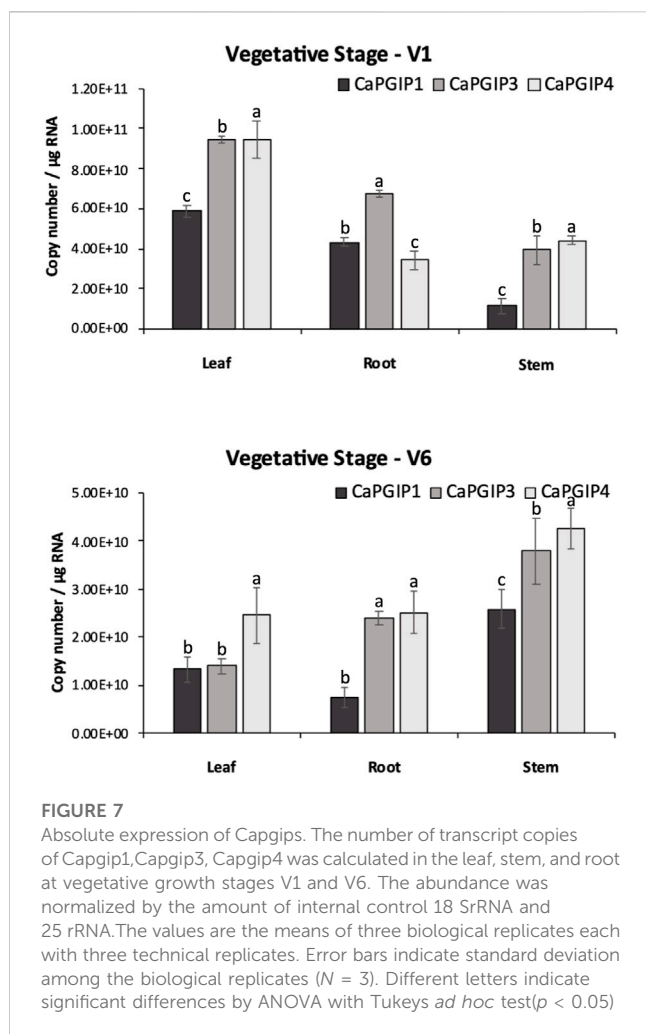
4 Discussion

Several plant species have PGIP-encoding small gene families; these multigene families encode proteins with similar LRR domains but different PG-inhibitory capabilities (D'Ovidio et al., 2006). This study demonstrated that the chickpea genome has a PGIP family of potentially three functional genes (*CaPGIP1*, *CaPGIP3*, and *CaPGIP4*) that are present on two chromosomes. Soybean is the only other legume where PGIPs are located on multiple chromosomes (D'Ovidio et al., 2006). Like the genomic distribution of *CaPGIPs*, legume PGIPs are distributed across a broader genomic region, as observed in soybean, where *GmPGIP1*, *GmPGIP2*, and *GmPGIP5* are located on chromosome 5, and span a ~ 19.5 kbp region, while *GmPGIP3*, *GmPGIP4*, and *GmPGIP7*, present on chromosome 8, span a ~ 21 kbp region. In common bean, *PvPGIP1*, *PvPGIP2*, *PvPGIP3*, and *PvPGIP4* span an area of ~ 50 kbp on chromosome 2 (D'Ovidio et al., 2004; D'Ovidio et al., 2006).

So far, over eighteen legume PGIPs have been either computationally or biochemically characterized in 9 species. The reported sequence lengths of these legume PGIPs ranged from

321 amino acids for MtPGIP2 to 342 amino acids for MtPGIP1. All these PGIPs contain signal peptides, with GmPGIP4 having the shortest with 17 amino acids and PvPGIP2 having the longest with 29 amino acids. Protein molecular mass (kDa) ranged from 35.92 for MtPGIP2 to 38.17 for MtPGIP1. And the isoelectric point (pI) ranged from 6.79 for PsPGIP1 to 9.48 for PvPGIP4. Except for MsPGIP2 of alfalfa, all other legume PGIPs had 10 LRRs and all were intronless. Only MsPGIP2's genomic sequence contains 9 LRRs and a 154-bp intron. *CaPGIP1*, *CaPGIP3*, and *CaPGIP4* exhibit very similar above-mentioned characteristics, with the exception that *CaPGIP1* has the longest signal peptide among the known legume PGIPs so far, with 36 amino acids. Because of selection for transcription efficiency, conserved genes with relatively high levels of expression tend to lose introns (Zou et al., 2011). Intronless genes such as PGIP may play key roles in plant growth, development, or response to biotic or abiotic stresses (Liu et al., 2021).

Like many PGIPs, the central domain of *CaPGIP1*, *CaPGIP3*, and *CaPGIP4* is comprised of 10 imperfect leucine-rich repeats (LRRs). Each is about 24 amino acids long and perfectly matches the extracellular LRR (eLRRs) consensus sequence xxLxLxx.NxLx. GxIPxxLxxLxxL (Di Matteo et al., 2003). eLRRs are important in plant defense as they function as receptor-like proteins or receptor-like kinases to recognize diverse pathogen molecules, and plant hormones (van der Hoorn et al., 2005). The presence of xxLxLxxNxL core consensus in eLRRs is responsible for the β -sheet structure formation (Zambounis et al., 2012). The LRRs are arranged tandemly in the PGIP's distinctive curved and elongated shape. The β -sheets are parallelly organized on the



inner side of the protein to form the concave face, while the 3_{10} -helices are parallelly organized on the outer side to make the convex face (Kobe and Deisenhofer, 1993). Solvent-exposed residues on the concave-sheet surface bind to pathogen molecules. CaPGIPs, have two types of β -sheets ($\beta 1$ and $\beta 2$) and 3_{10} -helices. All LRRs of plant defense proteins have $\beta 1$ sheet; however, the presence of $\beta 2$ sheets is unique and seen only in PGIPs (Di Matteo et al., 2003). CaPGIPs are predicted to have several N-glycosylation on these β -sheets which are vital for ligand binding for disease resistance, and the heterogeneity in β -sheet residues or the glycosylation patterns contributes to the varying recognition specificities of LRR proteins (Ramanathan et al., 1997; van der Hoorn et al., 2005). The deduced CaPGIP proteins also contain conserved cysteine residues that form disulfide bridges crucial for the maintenance of secondary structures in PGIP (Veronico et al., 2011).

Previous research indicated that truncated variants of functioning NBS-LRR genes can be found within 100 kb of fully functional NBS-LRR genes. These truncated genes are often pseudogenes because of alternative splicing. These pseudogenes have large deletions due to various transposition events (Marone et al., 2013). CaPGIP2 is found within 30 kb of CaPGIP1. CaPGIP2's C-terminal end matches CaPGIP1 perfectly, implying that the majority of the central LRR

domain and N-terminus might be deleted. Also, the presence of a nearly 10-kbp intron is atypical for PGIPs and most functional genes. Since it also lacks a fundamental signal peptide, it is classified as a non-secretory protein. For these reasons, CaPGIP2 was not subjected to investigation in this study beyond the subcellular localization analysis.

Alignment of CaPGIP amino acid sequences with amino acid sequences from various other legume PGIPs revealed that CaPGIPs are similar to those of other characterized PGIPs. Interestingly, alignment also indicated that CaPGIPs along with other legume PGIPs are highly conserved at the $\beta 2$ -sheet sites, which may be because $\beta 2$ -sheet are present only in PGIPs and are absent in other LRR proteins. Even though there is a higher level of similarity in the β sheet regions, there are also many variable portions present in both β -sheets. This variability is most likely responsible for the presence of multiple recognition specificities to target broader pathogen PGs (De Lorenzo et al., 2001; Matsushima and Miyashita, 2012). As per the phylogenetic analysis, CaPGIPs have a high degree of similarity with PGIPs from different plant sources. CaPGIP1 shared up to 70% similarity with *P. sativum* PsPGIP1 (AJI49705.1), and 69% with *G. max* GmPGIP3 (NP 001304551.2). It has been previously reported that PsPGIP1 and GmPGIP3 potentially inhibit several pathogens. For instance, PsPGIP1 has been identified as a possible defense factor against the pea-cyst nematode *Heterodera goettingiana* (Veronico et al., 2011). Encoded protein products of GmPGIP3 inhibited PGs from *Sclerotinia*, *Fusarium*, and *Botrytis* (D'Ovidio et al., 2006; Wang et al., 2015). Transgenic wheat expressing GmPGIP3 also showed enhanced resistance to *Gaeumannomyces graminis* var. *tritici*, and *Bipolaris sorokiniana* (D'Ovidio et al., 2006; Wang et al., 2015). Due to CaPGIP1's high sequence similarity to PsPGIP1 and GmPGIP3, it is likely to have similar functions and be engaged in nematode or fungal disease inhibition.

CaPGIP3 is more similar to the two tightly linked PGIPs of *V. radiata*, with 64% and 69% similarity to VrPGIP1 and VrPGIP2, respectively. Both VrPGIPs are known to provide resistance to bruchids (*Callosobruchus* spp) (Kaewwongwal et al., 2017), and CaPGIP3 may play a similar role in chickpeas against bruchids. CaPGIP4 had higher similarity to tree fruit PGIPs, with 72% similarity to *Malus domestica* MdPGIP1 (AAB19212.1) and 71% similarity to *Pyrus communis* PcPGIP (AAA33865.1). MdPGIP1 protein inhibited PG production in *Colletotrichum lupini* and *Aspergillus niger* (Oelofse et al., 2006). Unlike CaPGIP1 and CaPGIP3, CaPGIP4 may only be effective against pathogenic fungi.

The TATA box and CAAT box motifs are widely found in functional gene's promoter and enhancer regions (Xue, 2002; Svensson et al., 2006). TATA boxes function as a motif for recruiting transcription initiation machinery and RNA polymerase II, while CAAT boxes improve protein binding (Joubert et al., 2013; Liao et al., 2015). TATA and CAAT boxes are conserved eukaryotic cis-elements that are found in many plant gene promoters, including PGIPs. CaPGIP1, CaPGIP3, and CaPGIP4 all have several TATA and CAAT boxes upstream of the start codon ATG, indicating that they are functioning genes. Plant PGIPs are typically expressed after pathogen infection and wounding response (Kalunke et al., 2015), hence the presence of multiple pathogenicity-related and wounding motifs in the CaPGIP1, CaPGIP3, and CaPGIP4 promoter region. Apart from pathogens, PGIPs are triggered by phytohormone treatment in several plant species (Stotz et al., 1993; Yao et al., 1999;

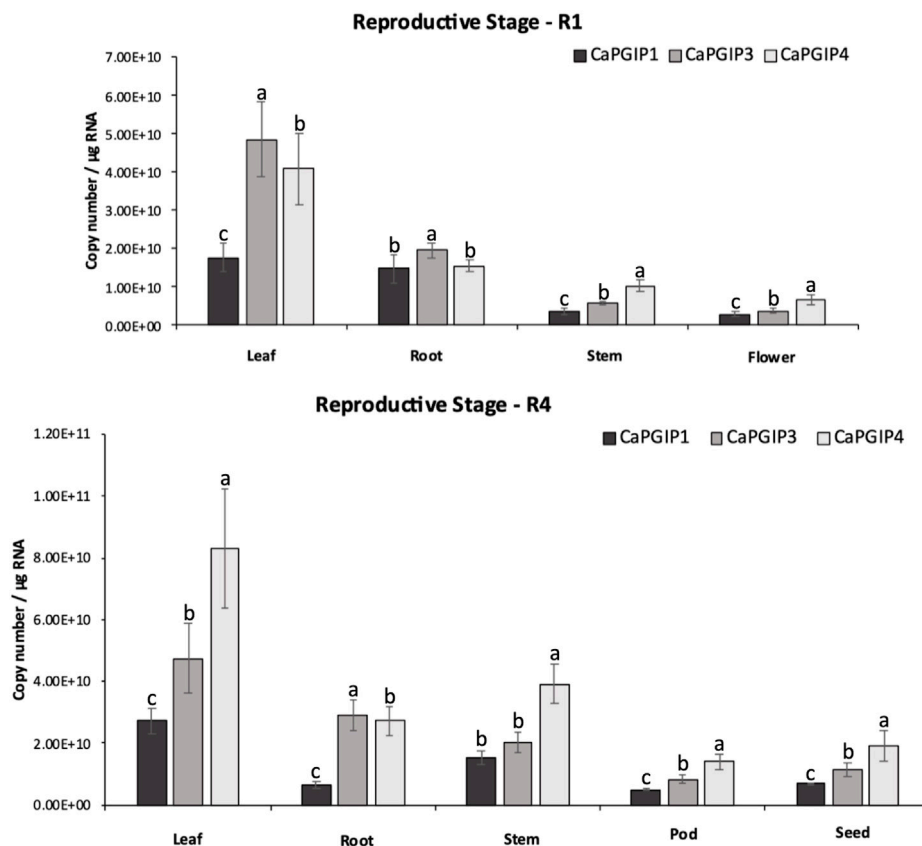


FIGURE 8

Absolute expression of *Capgips*. The numbers of transcript copies of *Capgip1*, *Capgip3*, and *Capgip4* were calculated in the leaf, stem, root and flower at reproductive growth stage R1, and leaf, stem, root, pod and seed at reproductive growth stage R4. The abundance was normalized by the amount of internal control 18 SrRNA and 25 rRNA. The value are the means of three biological replicates each with three technical replicates. Error bars indicate standard deviation among the biological replicates ($N = 3$) Different letters indicate significant differences by ANOVA with Tukeys *ad hoc* test ($p < 0.005$).

Ferrari et al., 2003). *PGIP* expression in rice, alfalfa, and pepper is induced by abscisic treatment. *PGIPs* from rapeseed, rice, barrel clover, and pepper are triggered by jasmonic acid and salicylic acid (Song and Nam, 2005; Janni et al., 2006; Hegedus et al., 2008; Lu et al., 2012; Wang et al., 2013). Rice *PGIPs* are induced by gibberellic acid treatment (Janni et al., 2006; Lu et al., 2012). The presence of multiple cis-acting elements in *CaPGIP1*, *CaPGIP3*, and *CaPGIP4* that regulate abscisic acid, gibberellic acid, jasmonic acid, and salicylic acid pathways suggests that they may play comparable roles to those seen in many plant *PGIPs*. Like other families of defense-related genes, *PGIPs* demonstrate tissue-specific activity. The grapevine *VvPGIP1* gene is only expressed in roots and ripening berries, and its expression is developmentally regulated (Joubert et al., 2006). Several cis-acting elements that influence tissue-specific responses, particularly root-specific responses, were found in all *CaPGIPs*, indicating a role in plant development or resistance to pathogens that enter the plant system through the roots. *Brassica juncea PGIPs* are associated with high temperature and drought stresses (Bhardwaj et al., 2015), while Arabidopsis *AtPGIP1*, and apple's *MdPGIP1*, are induced in response to cold stress (Kalunke et al., 2015). The presence of regulatory elements associated with drought, dehydration, water, high light, and low-temperature stress in the *CaPGIP* promoter suggests that they may play a role in plant stress.

Bioinformatic analysis and subcellular localization confirmed that *CaPGIP1*, *CaPGIP3*, and *CaPGIP4* are secretory proteins, and they are located in the plasma membrane or the cell wall. DeepLoc-1.0 analysis indicated *CaPGIP2* might be found in the mitochondrion or the cytoplasm, and localization experiments revealed *CaPGIP2* was found in the cytoplasm and endoplasmic reticulum. The locations of *CaPGIPs* within the plasma membrane/cell wall were consistent with the locations of other legume *PGIPs*. Localization on the plasma membrane is crucial as these proteins play a role in defense responses as cell surface receptors to detect pathogen PGs in the apoplast (Rodriguez-Palenzuela et al., 1991). *CaPGIP2* localization to the cytoplasm and endoplasmic reticulum might be because of the lack of signal peptide in *CaPGIP2* and suggests it might not be involved with the PG interaction in the apoplast.

Several studies investigated *PGIP* gene expression in response to external stimuli. On the other hand, analyzing gene expression in the absence of external stimuli or treatment enables the correlation of the expression of different *PGIP* genes within a crop. Furthermore, pathogens can infect plants at any stage of their life cycle, and *PGIP* gene families, like other defense-related gene families, have been demonstrated to exhibit variable expression patterns (Kalunke et al.,

2015). Because *PGIP* genes exhibit functional redundancy and sub-functionalization at the protein level (De Lorenzo et al., 2001; Federici et al., 2006), tissue-specific expression of *PGIP* genes is feasible, allowing them to respond more effectively to a variety of environmental stimuli (D'Ovidio et al., 2004). In terms of pathogen PG specificity, plant *PGIP* genes can express at higher levels in distinct growth stages and tissues that correspond to pathogen infection (Cantu et al., 2008). Absolute expression analysis showed *CaPGIP* genes has higher expression in leaf tissue and the least in pod tissues. *B. vulgaris*'s *BvPGIP* genes were reported to be highly expressed in roots in comparison to leaf tissue during normal growth and development (Li and Smigocki, 2016). *Carica papaya*'s *CpPGIP4* and *CpPGIP6* genes were shown to be ubiquitously expressed in root, stem, leaf, seed fruit pulp and peel. However, the *CpPGIP* gene transcripts were most abundant in fruit pulp and peel and decreased during ripening (Broetto et al., 2015). Tissue-specific differences have been reported in apples, where higher transcript abundance was in leaves and fruit, least in the stem (Zhang et al., 2010). In blackberry *PGIP* gene expression was more abundant in young leaves and fruit compared to old leaves and ripe fruit (Hu et al., 2012). In raspberries, *PGIP* transcripts were detected in fruit but not in flowers (Johnston et al., 1993). The varying expression levels of *CaPGIP* genes in different tissues at different growth stages indicates that *CaPGIP* genes might respond to different external stimuli.

5 Conclusion

In conclusion, this study is the first to characterize chickpea PGIPs. Two additional PGIPs on chromosome 3, *CaPGIP3* and *CaPGIP4* were identified in addition to the previously reported *CaPGIP1* and *CaPGIP2* on chromosome 6 and this necessitated modifying the genomic organization of *CaPGIPs*. *CaPGIP1*, *CaPGIP3*, and *CaPGIP4* displayed a typical PGIP sequence identity with an N-terminal domain, a central LRR domain with ten imperfect LRRs, and a C-terminal domain. Multiple sequence alignment shows that *CaPGIP* amino acid sequences are highly similar to those of other described legumes. The phylogenetic study of *CaPGIPs* indicated that *CaPGIP1* and *CaPGIP3* are similar to legume PGIPs, and *CaPGIP4* falls outside the legume PGIP cluster. *CaPGIP*'s promoter sequences harbor cis-elements that regulate response to various external stimuli. *CaPGIP1*, *CaPGIP3*, and *CaPGIP4* are localized to the cell wall or plasm membrane. Absolute quantification of the *CaPGIP* transcript levels under untreated conditions demonstrates that *CaPGIPs* have tissue-specific expression. Interestingly, *CaPGIP2* lacked most of the characteristics typical of a PGIP and warrants further investigations.

Data availability statement

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

Author contributions

VE and WC conceived the study and designed the experiments. VE performed the experiments. VE analyzed the data. VE wrote the manuscript with contributions from all co-authors. All authors contributed to the article and approved the submitted version.

Acknowledgments

We would like to thank Dan Mullendore of Franceschi Microscopy and Imaging Center, Washington State University, Pullman, WA for assistance with confocal microscopy. We would also like to thank Sheri McGrew, Grain Legume Genetics Physiology Research, USDA ARS, Pullman, WA for assistance with maintaining plants in the green house. We would also like to thank Tony Chen, Grain Legume Genetics Physiology Research, USDA ARS, Pullman, WA for assistance in laboratory work. We would also like to thank Derek Pouchnik and Weiwei Du of Laboratory for Biotechnology and Bioanalysis Washington State University, Pullman, WA for their assistance with plasmid and DNA sequencing. Financial support from the Indian Council for Agricultural Sciences (ICAR) is gratefully acknowledged.

Conflict of interest

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

Publisher's note

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/fgene.2023.1189329/full#supplementary-material>

SUPPLEMENTARY TABLE S1

Genes identified through a homology database search in *Cicer arietinum* genome that exhibited PGIP features.

SUPPLEMENTARY TABLE S2

Primers used in this study to isolate, clone, and for the RT-qPCR expression analysis of the *Capgips*.

SUPPLEMENTARY TABLE S3

Other putative *cis*-acting regulatory elements identified in the promoter regions of *Capgips*.

References

- Albersheim, P., and Anderson, A. J. (1971). Proteins from plant cell walls inhibit polygalacturonases secreted by plant pathogens. *Proc. Natl. Acad. Sci. U. S. A.* 68 (8), 1815–1819. doi:10.1073/pnas.68.8.1815
- Armenteros, J. J. A., Sønderby, C. K., Sønderby, S. K., Nielsen, H., and Winther, O. (2017). DeepLoc: Prediction of protein subcellular localization using deep learning. *Bioinforma. Oxf. Engl.* 33 (21), 3387–3395. doi:10.1093/bioinformatics/btx431
- Armenteros, J. J. A., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., et al. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.* 37, 420–423. doi:10.1038/s41587-019-0036-z
- Bhardwaj, A. R., Joshi, G., Kukreja, B., Malik, V., Arora, P., Pandey, R., et al. (2015). Global insights into high temperature and drought stress regulated genes by RNA-Seq in economically important oilseed crop *Brassica juncea*. *BMC plant Biol.* 15, 9. doi:10.1186/s12870-014-0405-1
- Bretag, T., and Horsham, H. K. (2004). Sclerotinia of chickpeas. *Agric. notes*.
- Broetto, S. G., Fabi, J. P., and Nascimento, J. R. O. do. (2015). Cloning and expression analysis of two putative papaya genes encoding polygalacturonase-inhibiting proteins. *Postharvest Biol. Technol.* 104, 48–56. doi:10.1016/j.postharvbio.2015.03.002
- Bustin, S. A., Benes, V., Garson, J. A., Hellemans, J., Huggett, J., Kubista, M., et al. (2009). The MIQE guidelines: Minimum information for publication of quantitative real-time PCR experiments. *Clin. Chem.* 55 (4), 611–622. doi:10.1373/clinchem.2008.112797
- Cantu, D., Vicente, A. R., Greve, L. C., Dewey, F. M., Bennett, A. B., Labavitch, J. M., et al. (2008). The intersection between cell wall disassembly, ripening, and fruit susceptibility to Botrytis cinerea. *Proc. Natl. Acad. Sci. U. S. A.* 105 (3), 859–864. doi:10.1073/pnas.0709813105
- D'Ovidio, R., Roberti, S., Di Giovanni, M., Capodicasa, C., Melaragni, M., Sella, L., et al. (2006). The characterization of the soybean polygalacturonase-inhibiting proteins (Pgp) gene family reveals that a single member is responsible for the activity detected in soybean tissues. *Planta* 224 (3), 633–645. doi:10.1007/s00425-006-0235-y
- Daher, F. B., and Braybrook, S. A. (2015). How to let go: Pectin and plant cell adhesion. *Front. plant Sci.* 6, 523. doi:10.3389/fpls.2015.00523
- De Lorenzo, G., D'Ovidio, R., and Cervone, F. (2001). The role of polygalacturonase-inhibiting proteins (PGIPs) in defense against pathogenic fungi. *Annu. Rev. Phytopathology* 39, 313–335. doi:10.1146/annurev.phyto.39.1.313
- Desiderio, A., Aracri, B., Leckie, F., Mattei, B., Salvi, G., Tigelaar, H., et al. (1997). Polygalacturonase-inhibiting proteins (PGIPs) with different specificities are expressed in *Phaseolus vulgaris*. *Phaseolus Vulgaris. Mol. Plant Microbe Interact.* 10, 852–860. doi:10.1094/MPMI.1997.10.7.852
- Di Giovanni, M., Cenci, A., Janni, M., and D'Ovidio, R. (2008). A LTR copia retrotransposon and Mutator transposons interrupt PGIP genes in cultivated and wild wheats. *Theor. Appl. Genet.* 116, 859–867. doi:10.1007/s00122-008-0719-1
- Di Matteo, A., Bonivento, D., Tsernoglou, D., Federici, L., and Cervone, F. (2006). Polygalacturonase-inhibiting protein (PGIP) in plant defense: A structural view. *Phytochemistry* 67 (6), 528–533. doi:10.1016/j.phytochem.2005.12.025
- Di Matteo, A., Federici, L., Mattei, B., Salvi, G., Johnson, K. A., Savino, C., et al. (2003). The crystal structure of polygalacturonase-inhibiting protein (PGIP), a leucine-rich repeat protein involved in plant defense. *Proc. Natl. Acad. Sci. U. S. A.* 100 (17), 10124–10128. doi:10.1073/pnas.1733690100
- D'Ovidio, R., Mattei, B., Roberti, S., and Bellincampi, D. (2004). Polygalacturonases, polygalacturonase-inhibiting proteins and pectic oligomers in plant-pathogen interactions. *Biochimica Biophysica Acta (BBA) - Proteins Proteomics* 1696 (2), 237–244. doi:10.1016/j.bbapap.2003.08.012
- Farina, A., Rocchi, V., Janni, M., Benedettelli, S., De Lorenzo, G., and D'Ovidio, R. (2009). The bean polygalacturonase-inhibiting protein 2 (PvPGIP2) is highly conserved in common bean (*Phaseolus vulgaris* L.) germplasm and related species. *Theor. Appl. Genet.* 118 (7), 1371–1379. doi:10.1007/s00122-009-0987-4
- Federici, L., Di Matteo, A., Fernandez-Recio, J., Tsernoglou, D., and Cervone, F. (2006). Polygalacturonase inhibiting proteins: Players in plant innate immunity? *Trends plant Sci.* 11 (2), 65–70. doi:10.1016/j.tplants.2005.12.005
- Ferrari, S., Vairo, D., Ausubel, F. M., Cervone, F., and De Lorenzo, G. (2003). Tandemly duplicated Arabidopsis genes that encode polygalacturonase-inhibiting proteins are regulated coordinately by different signal transduction pathways in response to fungal infection. *Plant Cell* 15 (1), 93–106. doi:10.1105/tpc.005165
- Forlenza, M., Kaiser, T., Savelkoul, H. F. J., and Wiegertjes, G. F. (2012). In *The use of real-time quantitative PCR for the analysis of cytokine mRNA levels BT* in cytokine protocols. Editor M. De Ley (Totowa, NJ: Humana Press), 7–23. doi:10.1007/978-1-61779-439-1_2
- Frati, F., Galletti, R., De Lorenzo, G., Salarino, G., and Conti, E. (2006). Activity of endopolygalacturonases in mirid bugs (*Heteroptera: Miridae*) and their inhibition by plant cell wall proteins (PGIPs). *Eur. J. Entomology* 103, 515–522. doi:10.14411/eje.2006.067
- Gao, J., Gui, Z., Wang, Y., Liu, H., Xin, N., Zhang, X., et al. (2015). Polymorphism of polygalacturonase-inhibiting protein genes by fluorescence-based PCR single-strand conformation profiling and sequencing in autotetraploid alfalfa (*Medicago sativa*) populations. *Euphytica* 206, 567–577. doi:10.1007/s10681-015-1455-7
- Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., et al. (2005). In *Protein identification and analysis tools on the ExPASy server* in the proteomics protocols handbook. Editor J. M. Walker Totowa, NJ: Humana Press Inc, 571–608. doi:10.1385/1-59259-890-0:571
- Ge, T., Huang, X., and Xie, R. (2019). Recent advances in polygalacturonase gene in fruit tree species. *Plant Physiology J.* 55, 1075–1088.
- Gupta, R., and Brunak, S. (2002). Prediction of glycosylation across the human proteome and the correlation to protein function. *Pac. Symposium Biocomput.*, 310–322.
- Haeger, W., Wielsch, N., Shin, N. R., Gebauer-Jung, S., Pauchet, Y., and Kirsch, R. (2021). New players in the interaction between beetle polygalacturonases and plant polygalacturonase-inhibiting proteins: Insights from proteomics and gene expression analyses. *Front. plant Sci.* 12, 660430. doi:10.3389/fpls.2021.660430
- Hamera, S., Mural, R. V., Liu, Y., and Zeng, L. (2014). The tomato ubiquitin-conjugating enzyme variant Suv, but not SLUev1C and SLUev1D regulates Fend-mediated programmed cell death in Nicotiana benthamiana. *Plant Signal. Behav.* 9 (10), e973814. doi:10.4161/15592324.2014.973814
- Hegedus, D. D., Li, R., Buchwaldt, L., Parkin, I., Whitwill, S., Coutu, C., et al. (2008). Brassica napus possesses an expanded set of polygalacturonase inhibitor protein genes that are differentially regulated in response to Sclerotinia sclerotiorum infection, wounding and defense hormone treatment. *Planta* 228 (2), 241–253. doi:10.1007/s00425-008-0733-1
- Higo, K., Ugawa, Y., Iwamoto, M., and Korenaga, T. (1999). Plant cis-acting regulatory DNA elements (PLACE) database. *Nucleic Acids Res.* 27 (1), 297–300. doi:10.1093/nar/27.1.297
- Hou, W., Mu, J., Li, A., Wang, H., and Kong, L. (2015). Identification of a wheat polygalacturonase-inhibiting protein involved in Fusarium head blight resistance. *Eur. J. Plant Pathology* 141 (4), 731–745. doi:10.1007/s10658-014-0574-7
- Hu, D., Dai, R., Wang, Y., Zhang, Y., Liu, Z., Fanf, R., et al. (2012). Molecular cloning, sequence analysis, and expression of the polygalacturonase-inhibiting protein (PGIP) gene in mulberry. *Plant Mol. Biol. Rep.* 30, 176–186. doi:10.1007/s11105-011-0324-3
- Hwang, B. H., Bae, H., Lim, H. S., Kim, K. B., Kim, S. J., Im, M. H., et al. (2010). Overexpression of polygalacturonase-inhibiting protein 2 (PGIP2) of Chinese cabbage (*Brassica rapa* ssp. *pekinensis*) increased resistance to the bacterial pathogen *Pectobacterium carotovorum* ssp. *Carotovorum*. *Plant Cell, Tissue, Organ Cult.* 103, 293–305. doi:10.1007/s11240-010-9779-4
- Janni, M., Di Giovanni, M., Roberti, S., Capodicasa, C., and D'Ovidio, R. (2006). Characterization of expressed Pgp genes in rice and wheat reveals similar extent of sequence variation to dicot PGIPs and identifies an active PGIP lacking an entire LRR repeat. *Theor. Appl. Genet.* 113 (7), 1233–1245. doi:10.1007/s00122-006-0378-z
- Jiménez-Fernández, D., Montes-Borrego, M., Jiménez-Díaz, R. M., Navas-Cortés, J. A., and Landa, B. B. (2011). In planta and soil quantification of *Fusarium oxysporum* f. sp. *ciceris* and evaluation of *Fusarium* wilt resistance in chickpea with a newly developed quantitative polymerase chain reaction assay. *Phytopathology* 101 (2), 250–262. doi:10.1094/phyto-07-10-0190
- Johnston, D. J., Ramanathan, V., and Williamson, B. (1993). A protein from immature raspberry fruits which inhibits endopolygalacturonases from *Botrytis cinerea* and other micro-organisms. *J. Exp. Bot.* 44 (5), 971–976. doi:10.1093/jxb/44.5.971
- Joubert, D. A., de Lorenzo, G., and Vivier, M. A. (2013). Regulation of the grapevine polygalacturonase-inhibiting protein encoding gene: Expression pattern, induction profile and promoter analysis. *J. Plant Res.* 126 (2), 267–281. doi:10.1007/s10265-012-0515-5
- Joubert, D. A., Slaughter, A. R., Kemp, G., Becker, J. V. W., Krooshof, G. H., Bergmann, C., et al. (2006). The grapevine polygalacturonase-inhibiting protein (VvPGIP1) reduces Botrytis cinerea susceptibility in transgenic tobacco and differentially inhibits fungal polygalacturonases. *Transgenic Res.* 15 (6), 687–702. doi:10.1007/s11248-006-9019-1
- Jukanti, A. K., Gaur, P. M., Gowda, C. L., and Chibbar, R. N. (2012). Nutritional quality and health benefits of chickpea (*cicer arietinum* L.) A review. *Br. J. Nutr.* 108, S11–S26. doi:10.1017/S0007114512000797
- Kaewwongwal, A., Chen, J., Somta, P., Kongjaimun, A., Yimram, T., Chen, X., et al. (2017). Novel alleles of two tightly linked genes encoding polygalacturonase-inhibiting proteins (VrPGIP1 and VrPGIP2) associated with the *Br* locus that confer bruchid (*callosobruchus* spp) resistance to mungbean (*Vigna radiata*) accession V2709. *Front. plant Sci.* 8, 1692. doi:10.3389/fpls.2017.01692
- Kalunke, R. M., Cenci, A., Volpi, C., O'Sullivan, D. M., Sella, L., Favaron, F., et al. (2014). The PGIP family in soybean and three other legume species: Evidence for a third-and-death model of evolution. *BMC Plant Biol.* 14, 189. doi:10.1186/s12870-014-0189-3
- Kalunke, R. M., Janni, M., Sella, L., David, P., Geffroy, V., Favaron, F., et al. (2011). Transcript analysis of the bean polygalacturonase inhibiting protein gene family reveals that PVPGP2 is expressed in the whole plant and is strongly induced by pathogen infection. *J. Plant Pathology* 93 (1), 141–148.
- Kalunke, R. M., Tundo, S., Benedetti, M., Cervone, F., De Lorenzo, G., and D'Ovidio, R. (2015). An update on polygalacturonase-inhibiting protein (PGIP), a leucine-rich repeat protein that protects crop plants against pathogens. *Front. Plant Sci.* 6, 146. doi:10.3389/fpls.2015.00146

- Khan, R. A., Bhat, A. T., and Kumar, K. (2012). Management of chickpea (*cicer arietinum* L) dry root rot caused by rhizoctonia bataticola (taub) butler. *Int. J. Res. Pharm. Biomed. Sci.* 3 (4), 1539–1548.
- Knights, E. J., and Hobson, K. B. (2016). Chickpea overview. Retrieved from <https://www.sciencedirect.com/science/article/pii/B9780081005965000354>.
- Kobe, B., and Deisenhofer, J. (1993). Crystal structure of porcine ribonuclease inhibitor, a protein with leucine-rich repeats. *Nature* 366 (6457), 751–756. doi:10.1038/366751a0
- Kubicek, C. P., Starr, T. L., and Glass, N. L. (2014). Plant cell wall-degrading enzymes and their secretion in plant-pathogenic fungi. *Annu. Rev. phytopathology* 52, 427–451. doi:10.1146/annurev-phyto-102313-045831
- Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). Mega X: Molecular evolutionary Genetics analysis across computing platforms. *Mol. Biol. Evol.* 35 (6), 1547–1549. doi:10.1093/molbev/msy096
- Leticun, I., and Bork, P. (2021). Interactive tree of life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic acids Res.* 49 (W1), W293–W296. doi:10.1093/nar/gkab301
- Li, H., and Smigocki, A. C. (2016). Wound induced Beta vulgaris polygalacturonase-inhibiting protein genes encode a longer leucine-rich repeat domain and inhibit fungal polygalacturonases. *Physiological Mol. Plant Pathology* 96, 8–18. doi:10.1016/j.pmpp.2016.06.004
- Liao, Y. L., 1,2 Shen, Y. B., Chang, J., Zhang, W. W., Cheng, S. Y., and Xu, F. (2015). Isolation, expression, and promoter analysis of GbWRKY2: A novel transcription factor gene from ginkgo biloba. *Int. J. Genomics* 2015, 607185. doi:10.1155/2015/607185
- Liu, H., Lyu, H.-M., Zhu, K., Peer, Y. Van de., and Cheng, Z.-M. (2021). The emergence and evolution of intron-poor and intronless genes in intron-rich plant gene families. *Plant J.* 105, 1072–1082. doi:10.1111/tpj.15088
- Lu, L., Zhou, F., Zhou, Y., Fan, X., Ye, S., Wang, L., et al. (2012). Expression profile analysis of the polygalacturonase-inhibiting protein genes in rice and their responses to phytohormones and fungal infection. *Plant Cell Rep.* 31 (7), 1173–1187. doi:10.1007/s00299-012-1239-7
- Markel, S., Wise, K., McKay, K., Goswami, R., and Gudmestad, N. (2008). Ascochyta blight of chickpea. Retrieved from North Dakota State University website [https://www.ndsu.edu/pubweb/~goswami/Publications/Aschochyta%20Blight%20fact-sheet.pdf](http://efaidnbmnnnibpcajpcglclefindmkaj/https://www.ndsu.edu/pubweb/~goswami/Publications/Aschochyta%20Blight%20fact-sheet.pdf).
- Marone, D., Russo, M. A., Laidò, G., De Leonardi, A. M., and Mastrangelo, A. M. (2013). Plant nucleotide binding site-leucine-rich repeat (NBS-LRR) genes: Active guardians in host defense responses. *Int. J. Mol. Sci.* 14 (4), 7302–7326. doi:10.3390/ijms14047302
- Matsaunyan, L. B., Oelofse, D., and Dubery, I. A. (2015). *In silico* analysis of the polygalacturonase inhibiting protein 1 from apple, *Malus domestica*. *BMC Res. Notes* 8 (1), 76.
- Matsushima, N., and Miyashita, H. (2012). Leucine-rich repeat (LRR) domains containing intervening motifs in plants. *Biomolecules* 2, 288–311. doi:10.3390/biom2020288
- Mazur, S., Nawrocki, J., and Kućmierz, J. (2002). Fungal diseases of chickpea (*cicer arietinum* L) cultivated in the south region of Poland. *Plant Prot. Sci.* 38 (Special Issue 2), 332–335. doi:10.17221/10483-pps
- Merga, B., and Haji, J. (2019). Economic importance of chickpea: Production, value, and world trade. *Cogent Food & Agric.* 5 (1), 1615718. doi:10.1080/23311932.2019.1615718
- Mohnen, D. (1999). “3.15 biosynthesis of pectins and galactomannans,” in *Comprehensive natural products chemistry*, 497–527. doi:10.1016/b978-0-08-091283-7.00099-0
- Mojsov, K. D. (2016). “Chapter 16 - Aspergillus enzymes for food industries,” in *New and future developments in microbial Biotechnology and bioengineering*. Editor V. K. Gupta (Amsterdam: Elsevier), 215–222. doi:10.1016/B978-0-444-63505-1.00033-6
- Moore, M., Ryley, T., Knights, P., Nash, G., and Chiplin, G., and , Cumming. (2011). *Chickpeas—varietal selection, paddock planning and disease management in 2011. Northern region (Goondiwindi). GRDC Update Papers April 2011.*
- Oelofse, D., Dubery, I. A., Meyer, R., Arendse, M. S., Gazendam, I., and Berger, D. K. (2006). Apple polygalacturonase inhibiting protein1 expressed in transgenic tobacco inhibits polygalacturonases from fungal pathogens of apple and the anthracnose pathogen of lupins. *Phytochemistry* 67 (3), 255–263. doi:10.1016/j.phytochem.2005.10.029
- Pfaffl, M. W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic acids Res.* 29 (9), e45. doi:10.1093/nar/29.9.e45
- Ramanathan, V., Simpson, C. G., Thow, G., Iannetta, P. P. M., McNicol, R. J., and Williamson, B. (1997). cDNA cloning and expression of polygalacturonase inhibiting proteins (PGIPs) from red raspberry (*Rubus idaeus*). *J. Exp. Bot.* 48 (6), 1185–1193. doi:10.1093/jxb/48.6.1185
- Rodriguez-Palenzuela, P., Burr, T. J., and Collmer, A. (1991). Polygalacturonase is a virulence factor in *Agrobacterium tumefaciens* biovar 3. *J. Bacteriol.* 173 (20), 6547–6552. doi:10.1128/jb.173.20.6547-6552.1991
- SAS Institute Inc (2013). *SAS® 9.4 guide to software updates and product changes*. Cary, NC: SAS Institute Inc.
- Sharma, S., Upadhyaya, H. D., Roorkiwal, M., Varshney, R. K., and Gowda, C. L. L. (2013). “4 - chickpea,” in *Genetic and genomic resources of Grain legume improvement*. Editors M. Singh, H. D. Upadhyaya, and I. S. Bisht (Oxford: Elsevier), 81–111. doi:10.1016/B978-0-12-397935-3.00004-9
- Song, K. H., and Nam, Y. W. (2005). Genomic organization and differential expression of two polygalacturonase-inhibiting protein genes from *Medicago truncatula*. *J. Plant Biol.* 48, 467–478. doi:10.1007/BF03030589
- Stotz, H. U., Powell, A. L., Damon, S. E., Greve, L. C., Bennett, A. B., and Labavitch, J. M. (1993). Molecular characterization of a polygalacturonase inhibitor from *Pyrus communis* L. cv Bartlett. *Plant physiol.* 102 (1), 133–138. doi:10.1104/pp.102.1.133
- Svensson, J. T., Crosatti, C., Campoli, C., Bassi, R., Stanca, A. M., Close, T. J., et al. (2006). Transcriptome analysis of cold acclimation in barley albina and xantha mutants. *Plant physiol.* 141 (1), 257–270. doi:10.1104/pp.105.072645
- Toubart, P., Desiderio, A., Salvi, G., Cervone, F., Daroda, L., and De Lorenzo, G. (1992). Cloning and characterization of the gene encoding the endopolygalacturonase-inhibiting protein (PGIP) of *Phaseolus vulgaris* L. *Plant J. Cell Mol. Biol.* 2 (3), 367–373. doi:10.1046/j.1365-313x.1992.t01-35-00999.x
- Untergasser, A., Nijveen, A., Rao, X., Bisseling, T., Geurts, R., and Leunissen, J. A. M. (2007). Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res.* 35, W71–W74. doi:10.1093/nar/gkm306
- van der Hoorn, R. A., Wulff, B. B., Rivas, S., Durrant, M. C., van der Ploeg, A., de Wit, P. J., et al. (2005). Structure-function analysis of cf-9, a receptor-like protein with extracytoplasmic leucine-rich repeats. *plant Cell* 17 (3), 1000–1015. doi:10.1105/tpc.104.028118
- Veronico, P., Melillo, M. T., Saponaro, C., Leonetti, P., Picardi, E., and Jones, J. T. (2011). A polygalacturonase-inhibiting protein with a role in pea defence against the cyst nematode *Heterodera goettingiana*. *Mol. plant Pathol.* 12 (3), 275–287. doi:10.1111/j.1364-3703.2010.00671.x
- Voragen, A. G. J., Coenen, G. J., Verhoef, R. P., and Schols, H. A. (2009). Pectin, a versatile polysaccharide present in plant cell walls. *Struct. Chem.* 20, 263–275. doi:10.1007/s11224-009-9442-z
- Wang, A., Wei, X., Rong, W., Dang, L., Du, L. P., Qi, L., et al. (2015). GmPGIP3 enhanced resistance to both take-all and common root rot diseases in transgenic wheat. *Funct. Integr. genomics* 15 (3), 375–381. doi:10.1007/s10142-014-0428-6
- Wang, X., Zhu, X., Tooley, P., and Zhang, X. (2013). Cloning and functional analysis of three genes encoding polygalacturonase-inhibiting proteins from *Capsicum annuum* and transgenic *CaPGIP1* in tobacco in relation to increased resistance to two fungal pathogens. *Plant Mol. Biol.* 81, 379–400. doi:10.1007/s11103-013-0007-6
- Wang, Y., Zhang, P., Li, L., Li, D., Liang, Z., Cao, Y., et al. (2022). Proteomic analysis of alfalfa (*Medicago sativa* L) roots in response to rhizobium nodulation and salt stress. *Genes* 13 (11). doi:10.3390/genes13112004
- Wang, Z., Wan, L., Zhang, X., Xin, Q., Song, Y., Hong, D., et al. (2021). Interaction between *Brassica napus* polygalacturonase inhibition proteins and *Sclerotinia sclerotiorum* polygalacturonase: Implications for rapeseed resistance to fungal infection. *Planta* 253 (2), 34. doi:10.1007/s00425-020-03556-2
- Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., et al. (2018). SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res.* 46 (W1), W296–W303–W303. doi:10.1093/nar/gky427
- Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., and Barton, G. J. (2009). Jalview Version 2-A multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25, 1189–1191. doi:10.1093/bioinformatics/btp033
- West, V. P., Appiah, A. A., and Gow, A. R. N. (2003). Advances in research on oomycete root pathogens. *Physiological Mol. Plant Pathology* 62 (2), 99–113. doi:10.1016/s0885-5765(03)00044-4
- Wong, M. L., and Medrano, J. F. (2005). Real-time PCR for mRNA quantitation. *BioTechniques* 39 (1), 75–85. doi:10.2144/05391RV01
- Wunsch, M. (2014). Management of Ascochyta blight of chickpea. Retrieved from Montana State University website <http://pspp.msuxextension.org/documents/NDSUCHICKPEA%20Aschochyta%20mgmt.pdf>.
- Xue, G. P. (2002). Characterisation of the DNA-binding profile of barley HvCBF1 using an enzymatic method for rapid, quantitative and high-throughput analysis of the DNA-binding activity. *Nucleic acids Res.* 30 (15), e77. doi:10.1093/nar/gnf076
- Yao, C., Conway, W. S., Ren, R., Smith, D., Ross, G. S., and Sams, E. (1999). Gene encoding polygalacturonase inhibitor in apple fruit is developmentally regulated and activated by wounding and fungal infection. *Plant Mol. Biol.* 39, 1231–1241. Available at: <https://naldc.nal.usda.gov/download/31319/PDF>. doi:10.1023/a:1006155723059
- Zambounis, A., Elias, M., Sterck, L., Maumus, F., and Gachon, C. M. (2012). Highly dynamic exon shuffling in candidate pathogen receptors. what if Brown algae were capable of adaptive immunity? *Mol. Biol. Evol.* 29 (4), 1263–1276. doi:10.1093/molbev/msr296
- Zhang, Q., Yan, Q., Yuan, X., Lin, Y., Chen, J., Wu, R., et al. (2021). Two polygalacturonase-inhibiting proteins (VrPGIP) of *Vigna radiata* confer resistance to bruchids (*Callosobruchus spp.*). *J. plant physiology* 258–259, 153376. doi:10.1016/j.jplph.2021.153376
- Zhou, B., Mural, R. V., Chen, X., Oates, M. E., Connor, R. A., Martin, G. B., et al. (2017). A subset of ubiquitin-conjugating enzymes is essential for plant immunity. *Plant Physiol.* 173 (2), 1371–1390. doi:10.1104/pp.16.01190
- Zou, M., Guo, B., and He, S. (2011). The roles and evolutionary patterns of intronless genes in deuterostomes. *Int. J. Genomics* 2011, 680673. doi:10.1155/2011/680673



OPEN ACCESS

EDITED BY

Rupesh Deshmukh,
Central University of Haryana, India

REVIEWED BY

Balpreet Kaur Dhatt,
University of Nebraska-Lincoln,
United States
Jun Tang,
Jiangsu Academy of Agricultural
Sciences, China

*CORRESPONDENCE

Wuwei Ye,
✉ yew158@163.com

[†]These authors have contributed equally
to this work

RECEIVED 29 March 2023

ACCEPTED 15 May 2023

PUBLISHED 07 June 2023

CITATION

Lei Y, Cui Y, Cui R, Chen X, Wang J, Lu X,
Wang D, Wang S, Guo L, Zhang Y, Rui C,
Fan Y, Han M, Zhao L, Zhang H, Liu X,
Xu N, Wang J, Huang H, Feng X, Xi Y, Ni K,
Zhang M, Jiang T and Ye W (2023),
Characterization and gene expression
patterns analysis implies BSK family genes
respond to salinity stress in cotton.
Front. Genet. 14:1169104.
doi: 10.3389/fgene.2023.1169104

COPYRIGHT

© 2023 Lei, Cui, Cui, Chen, Wang, Lu,
Wang, Wang, Guo, Zhang, Rui, Fan, Han,
Zhao, Zhang, Liu, Xu, Wang, Huang, Feng,
Xi, Ni, Zhang, Jiang and Ye. This is an
open-access article distributed under the
terms of the [Creative Commons
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). 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 and gene expression patterns analysis implies BSK family genes respond to salinity stress in cotton

Yuqian Lei^{1,2†}, Yupeng Cui^{3†}, Ruifeng Cui^{3†}, Xiugui Chen^{2†},
Junjuan Wang², Xuke Lu², Delong Wang², Shuai Wang²,
Lixue Guo², Yuexin Zhang², Cun Rui², Yapeng Fan², Mingge Han²,
Lanjie Zhao², Hong Zhang², Xiaoyu Liu², Nan Xu², Jing Wang²,
Hui Huang², Xixian Feng², Yanlong Xi², Kesong Ni²,
Menghao Zhang^{1,2}, Tiantian Jiang² and Wuwei Ye^{1,2*}

¹Zhengzhou Research Base, State Key Laboratory of Cotton Biology, School of Agricultural Sciences, Zhengzhou University, Zhengzhou, Henan, China, ²State Key Laboratory of Cotton Biology, Institute of Cotton Research of Chinese Academy of Agricultural Sciences, Anyang, Henan, China, ³Anyang Institute of Technology, Anyang, Henan, China

Identification, evolution, and expression patterns of BSK (BR signaling kinase) family genes revealed that BSKs participated in the response of cotton to abiotic stress and maintained the growth of cotton in extreme environment. The steroidal hormone brassinosteroids (BR) play important roles in different plant biological processes. This study focused on BSK which were downstream regulatory element of BR, in order to help to decipher the functions of BSKs genes from cotton on growth development and responses to abiotic stresses and lean the evolutionary relationship of cotton BSKs. BSKs are a class of plant-specific receptor-like cytoplasmic kinases involved in BR signal transduction. In this study, bioinformatics methods were used to identify the cotton BSKs gene family at the cotton genome level, and the gene structure, promoter elements, protein structure and properties, gene expression patterns and candidate interacting proteins were analyzed. In the present study, a total of 152 BSKs were identified by a genome-wide search in four cotton species and other 11 plant species, and phylogenetic analysis revealed three evolutionary clades. It was identified that BSKs contain typical PKc and TPR domains, the N-terminus is composed of extended chains and helical structures. Cotton BSKs genes show different expression patterns in different tissues and organs. The gene promoter contains numerous *cis*-acting elements induced by hormones and abiotic stress, the hormone ABA and Cold-inducing related elements have the highest count, indicating that cotton BSK genes may be regulated by various hormones at different growth stages and involved in the response regulation of cotton to various stresses. The expression analysis of BSKs in cotton showed that the expression levels of *GhBSK06*, *GhBSK10*, *GhBSK21* and *GhBSK24* were significantly increased with salt-inducing. This study is helpful to analyze the function of cotton BSKs genes in growth and development and in response to stress.

KEYWORDS

BR, BSK, brassinosteroid-signaling kinase, abiotic stress, salt stress

1 Introduction

With a habit of sessile growth, plants need to evolve a variety of regulatory methods to adapt to changes in the environment and deal with stress. Brassinosteroids (BR) has been reported as an important plant hormone involved in plant cell proliferation, senescence, male sterility, inducing flowering and fruit ripening and resisting abiotic stress (Clouse and Sasse, 1998; Nakaya et al., 2002; Symons et al., 2006; Bajguz and Hayat, 2009; Li et al., 2010; Ye et al., 2010; Fàbregas and Caño-Delgado, 2014). In the BR signaling pathway model that has been explored and established, many regulatory elements play an important role. The upstream elements include BRI (BR insensitive), BKI (BRI kinase inhibitor 1), BAK (BRI associated receptor kinase) and BSK (BR signaling kinase), and the downstream elements are jointly regulated by BIN2 (BR insensitive 2), BSU (BRI1 suppressor), BZR (BR resistant transcription factor) and BES (BRI-EMS suppressor) (Choudhary et al., 2012).

BSK is a class of receptor-like cytoplasmic kinases involved in BR signaling transduction. BSK is a phosphorylated substrate of BRI1, and after activation, it transmits signals downstream and is involved in regulating the expression of BR responsive genes (He et al., 2000). BSKs belong to the RLCK-XII superfamily, and their encoded proteins contain two typical kinase domains, PKc (Putative kinase catalytic) and TPR (Tetra-tri-co peptide repeat domains) domains. PKc can play a role in cell division, proliferation, apoptosis and differentiation, and TPR can mediate the interaction between proteins.

As an important cash crop, cotton inevitably suffers from adversity during the growth process, which affects its yield and quality. Therefore, revealing the components of cotton BR signal and its transduction pathway is of great significance for the response to stress and the regulation of growth and development. There are 12 BSKs in the model plant *Arabidopsis thaliana* (Li et al., 2019). Except for the bsk3-1 mutant which significantly reduced sensitivity to BR, the other single mutants showed no obvious phenotypic changes (Tang et al., 2008). However, quadruple mutants bsk3, 4, 7, 8 and quintuple mutants bsk3, 4, 6, 7, 8 significantly reduced the sensitivity and growth phenotype of BR, indicating that plant BSKs have functional redundancy (Kim et al., 2009; Kim et al., 2011; Sreeramulu et al., 2013). BSKs regulate BR signaling via an allosteric mechanism constitutively inactive protein kinases (Grütter et al., 2013).

Recent studies have pointed out that *Arabidopsis* BSK1 is physiologically related to pathogen-related molecular patterns PAMP (Pathogen-associated molecular patterns) and FLS2 (Flagellin-sensitive 2), and participates in the positive regulation of PAMP-triggered immunity (PAMP-triggered immunity), preventing the spread of pathogen infection and improving plant disease resistance (Stout et al., 2006; Bent and Mackey, 2007; Schwessinger and Zipfel, 2008; Boller and Felix, 2009; Shi et al., 2013). So far, there have been no reports on the functional studies of cotton BSKs genes in hormone signal transduction, growth and development, and response to stress. The expression level of two BSKs in transgenic *ScALDH21* cotton was higher than that of non-transgenic cotton under salt stress (Yang et al., 2023). Overexpression of BSKs in crops such as rice and maize has been found to enhance salt tolerance (Li et al., 2022; Liu et al., 2022; Zhang et al., 2022). However, the detailed characterization of BSK family proteins and their important function in plants remains unclear.

In this study, bioinformatics methods were used to analyze the gene structure, evolution, and expression patterns of cotton BSKs gene family. It lays the foundation for the transduction pathway, thereby providing new genetic resources for cotton breeding.

2 Materials and methods

2.1 Data download

The genome annotation files, protein sequence data, the genome sequence information (fasta format file) and genome gene structure annotation information (gff3 format file) of the four cotton species *Gossypium arboreum* (CRI), *Gossypium raimondii* (JGI), *Gossypium hirsutum* (ZJU) and *Gossypium barbadense* (ZJU) were obtained from Cotton Functional Genomics Database (CottonFGD) (Zhu et al., 2017) (<http://www.cottonfgd.org/>). The latest Genome annotation files and protein sequence data of other seven species *Arabidopsis thaliana*, *TAIR10*, *Glycine max_v2.1*, *Oryza sativa*, *IRGSP-1.0*, *Populus trichocarpa*, *Pop_tri_v3*, *Theobroma cacao_20110822*, *Vitis vinifera.12X* and *Zea mays. B73_RefGen_v4* were provided by JGI (<https://gold.jgi.doe.gov/>).

These softwares were used: TBtools (<https://github.com/CJ-Chen/TBtools/releases>), BLAST+ (<https://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/>), HMMER3.0 (<http://www.hmmer.org/>) and MEGA7 (<https://www.megasoftware.net/>). The Hidden Markov Models (HMMs) profile of PF07714 (PKc) and PF07719 (TPR) were downloaded from Pfam (<https://pfam.xfam.org/>).

2.2 Identification of BSK family genes

The BSK family members of *Arabidopsis* have been identified, with a total of 12 members (Li et al., 2019). The protein sequences of the 12 AtBSK family members were downloaded from the Tair website (<https://www.arabidopsis.org/>) as seed sequences to perform multiple sequence alignments in the plant protein database with BLAST + to identify the BSK family members of *G. hirsutum*, *G. barbadense*, *G. arboreum*, *G. raimondii*, *G. max*, *O. sativa*, *P. trichocarpa*, *T. cacao*, *V. vinifera* and *Z. mays*. The e value was less than 1e-5. The protein sequences of the members were queried with HMMER3.0 to obtain the correct family members (Zhang et al., 2021). TBtools helped to rename all the identified genes according to the gene ID to facilitate subsequent analysis.

2.3 Phylogenetic analysis and sequences alignments

In order to explore the evolutionary relationship between BSK family members, it is necessary to submit the protein sequences to MEGA7 (Kumar et al., 2016) to calculate the optimal evolutionary model and to construct a phylogenetic tree using the maximum likelihood method based on this model. The detailed parameters were as follows: maximum likelihood, Bootstrap replications 500, Jones-taylor-Thornton (JTT) model, Gamma distributed with Invariant sites (G + I), No of Discrete Gamma Categories 5,

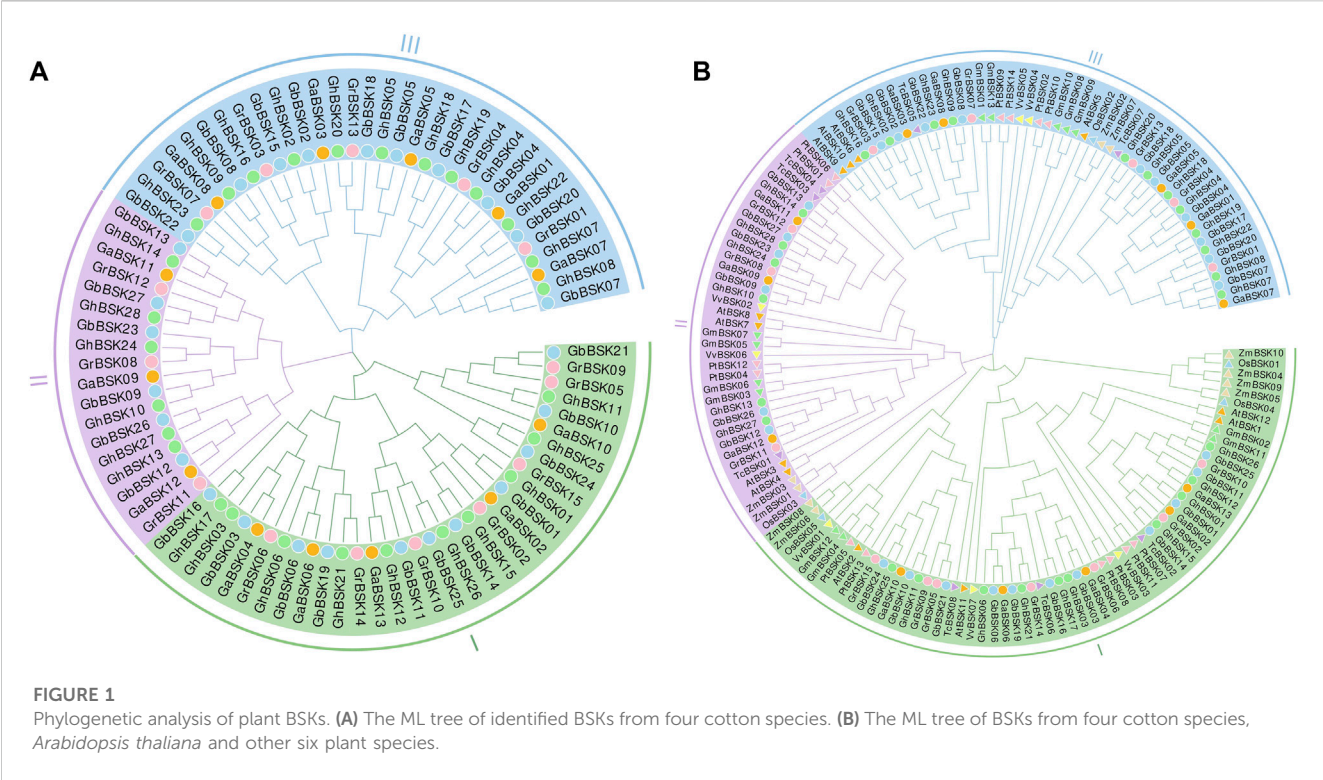


TABLE 1 Distribution of BSKs in different species.

Species	BSK I	BSK II	BSK III	Total
<i>G. hirsutum</i>	10	6	12	28
<i>G. barbadens</i>	11	6	10	27
<i>G. arboreum</i>	5	3	5	13
<i>G. raimondii</i>	7	3	5	15
<i>A. thaliana</i>	4	4	4	12
<i>G. max</i>	4	4	5	13
<i>O. sativa</i>	3	1	1	5
<i>P. trichocarp</i>	6	4	4	14
<i>T. cacao</i>	3	3	2	8
<i>V. vinifera</i>	3	2	2	7
<i>Z. mays</i>	6	2	2	10

Partial deletion 80 (Huang et al., 2022). The Evolview website (<https://www.evolgenius.info/evolview/>) was used for visual analysis of the evolutionary tree (Wang et al., 2021).

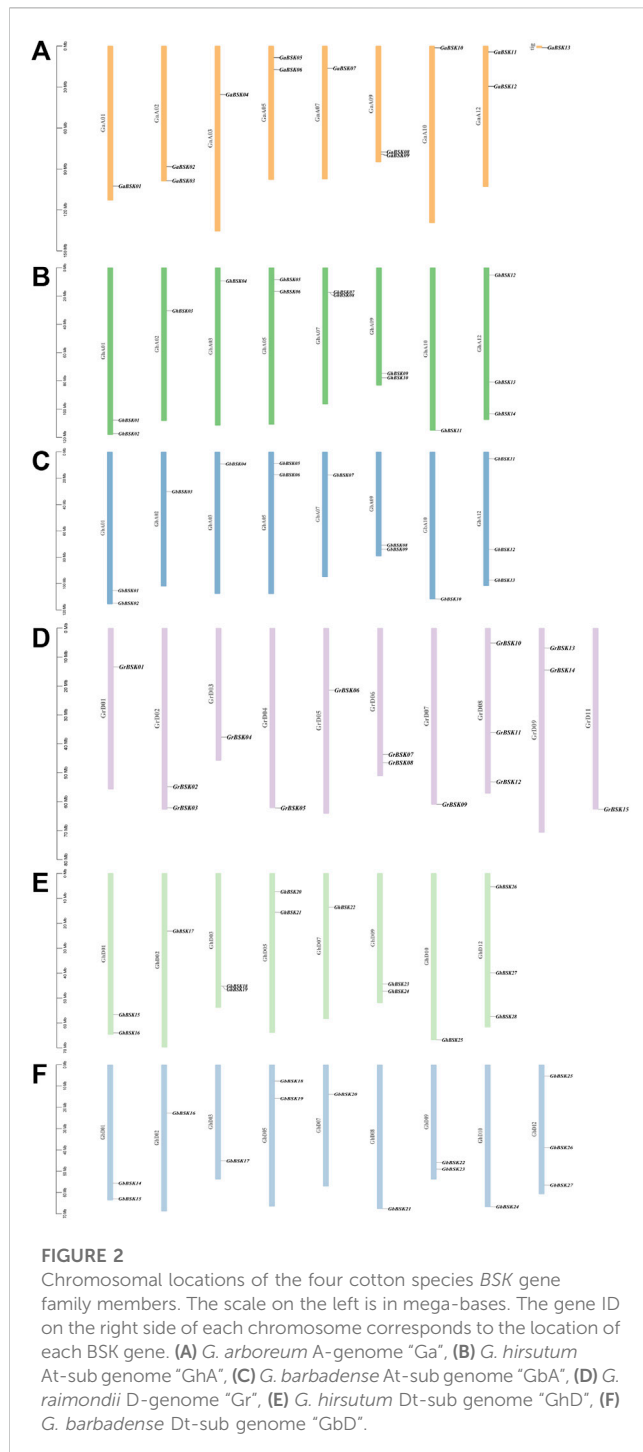
2.4 Conserved protein motifs, gene structure analysis and prediction of subcellular localization

The identified BSK protein sequences of four cotton species were submitted to the Motif Elicitation (MEME) (<https://meme-suite.org/>) online website to obtain the conserved protein motifs

TABLE 2 Number distribution of cotton BSKs in different chromosome.

Chromosomes	GhA	GhD	GbA	GbD	GaA	GrD
Chir 1	2	2	2	2	1	1
Chir 2	1	1	1	1	2	2
Chir 3	1	2	1	1	1	1
Chir 4						1
Chir 5	2	2	2	2	2	1
Chir 6						2
Chir 7	2	1	1	2	1	1
Chir 8						3
Chir 9	2	2	2	2	2	2
Chir 10	1	1	1	1	1	
Chir 11						1
Chir 12	3	3	3	3	2	
Chir 13						
Tig					1	
Total	14	14	13	14	13	15

and save the MAST xml file. The nwk file of cotton phylogenetic tree which was construct by MEGA7 was also needed. The MAST xml file, nwk file and gff3 genome file of cotton were placed to TBtools to gain a phylogenetic tree with conserved motifs and gene structures. BSK family protein sequences were submitted to



the website WoLF PSORT: Protein Subcellular Localization Prediction (<https://wolfpsort.hgc.jp/>) for subcellular localization prediction.

2.5 Collinearity analysis of *BSK* genes in four cotton species

The collinearity between the *BSK* duplicate gene pairs of the four cotton species was analyzed by MCScanX software on the four

cotton protein sequence data files, and the graphical results were visualized with TBtools software.

2.6 Calculation of selection pressure

The CDS sequences of the *BSK* genes in the four cotton species was downloaded from CottonFGD (<http://www.cottonfgd.org/>). The homologous gene pairs in the four cotton species from duplicated gene pairs were obtained by TBtools. In order to investigate the selection pressure, we calculated the ratio of the number of non-synonymous substitutions per non-synonymous site to the number of synonymous substitutions per synonymous site (Ka/Ks) for duplicated genes.

2.7 *Cis*-acting elements and expression pattern analysis

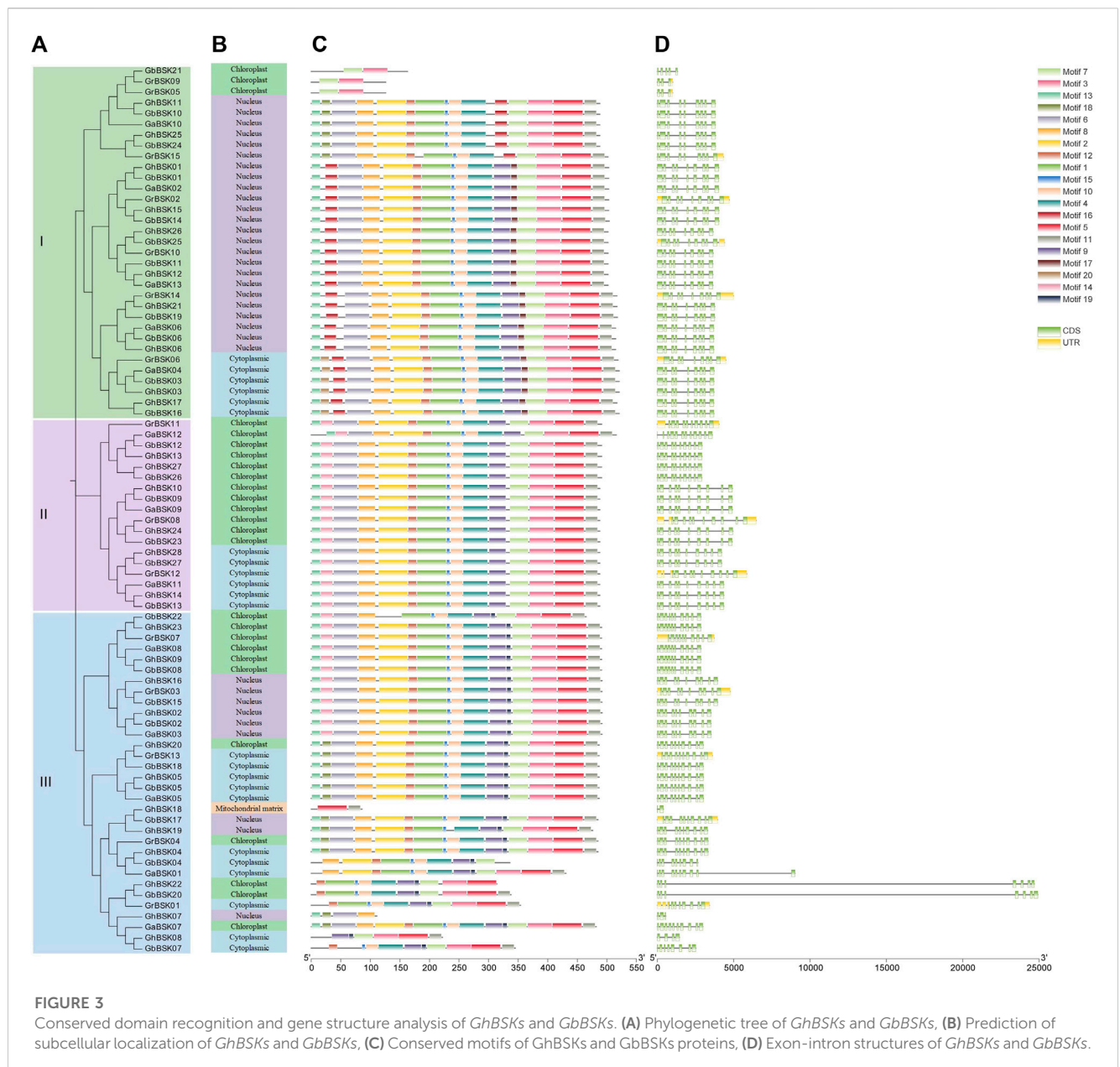
For the analysis of the promoter region of *BSK*, the upstream 2000 bp DNA sequence of *BSK* was downloaded from CottonFGD and uploaded to PlantCare website (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>) to predict the *cis*-acting elements (Wang et al., 2021). The nwk file of *BSK* phylogenetic tree, *cis*-acting element prediction results and CDS were submitted to TBtools for visual analysis.

For the analysis of the expression pattern (fragments per kilo base of exon per million mapped, FPKM) of *BSK* family genes, we downloaded RNA-Seq data from GRAND (Hu et al., 2019) (*Gossypium* Resource And Network Database) (<http://grand.cricaas.com.cn/home>) and draw heatmap based on the expression levels of *BSK* family genes via TBtools. The expression level change of *GhBSKs* under salt stress was shown in bar graph.

3 Results

3.1 Identification of *BSK* family members

After BLAST + sequence alignment, there are 28, 27, 13 and 15 members of *BSK* gene family in *G. hirsutum*, *G. barbadense*, *G. arboreum* and *G. raimondii*, a total of 83 were identified. The number of *BSK* members of tetraploid *G. hirsutum* and *G. barbadense* is twice that of diploid *G. arboreum* and *G. raimondii*. This became strong evidence that tetraploid cotton was an allotetraploid formed by hybridizing A-genome diploid and D-genome diploid. As for the other species, *G. max*, *O. sativa*, *P. trichocarpa*, *T. cacao*, *V. vinifera* and *Z. mays* were identified 13, 5, 14, 8, 7 and 10 *BSK* members. A total of 140 members were identified from the 10 species, and only one complete splice mutant was retained for each genomic locus. The total 140 *BSK* genes identified in these 10 species were sorted by gene ID and renamed with TBtools as *GhBSK01* - *GhBSK28*, *GbBSK01* - *GbBSK27*, *GaBSK01* - *GaBSK13*, *GrBSK01* - *GrBSK15*, *GmBSK01* - *GmBSK13*, *OsBSK01* - *OsBSK05*, *PtBSK01* - *PtBSK14*, *TcBSK01* - *TcBSK08*, *VvBSK01* - *VvBSK07* and



ZmBSK01 - *ZmBSK10* (Supplementary Table S1). The names of *AtBSK* family members were consistent with Tair website. The *BSKs* of tetraploid cotton has more *BSK* members than diploid plants such as rice and grapes, indicating that *BSK* has gene duplication in evolution.

3.2 Phylogenetic analysis and evolution analysis

The phylogenetic relationship between *BSK* proteins is shown in the phylogenetic tree shown in Figure 1. The phylogenetic tree shows that *BSKs* have three distinct evolutionary branches, classified as *BSK* I - *BSK* III. *ZmBSKs* and *OsBSKs* proteins were more closely related. The number of different classes of *BSKs* from different species were shown in Table 1. There were

more *BSK* I and *BSK* III present in cotton, twice as much as *BSK* II. *Arabidopsis thaliana*, *G. max*, *V. vinifera* and *T. cacao* have an even distribution. *O. sativa*, *P. trichocarpa* and *Z. mays* have more *BSK* I, but less *BSK* II and *BSK* III.

3.3 Chromosomal localization of four cotton species

Chromosome distribution maps show the distribution of cotton *BSK* members on chromosomes in detail, and counted the number of genes on each chromosome (Table 2). The distribution in *G. hirsutum*, *G. barbadense*, and *G. arboreum* is highly conserved, and the fourth, sixth, eighth (excepted *GbD*), 11th and 13th chromosomes in each group of chromosomes have no *BSK* distribution (Figure 2). While *G. raimondii* is different from the

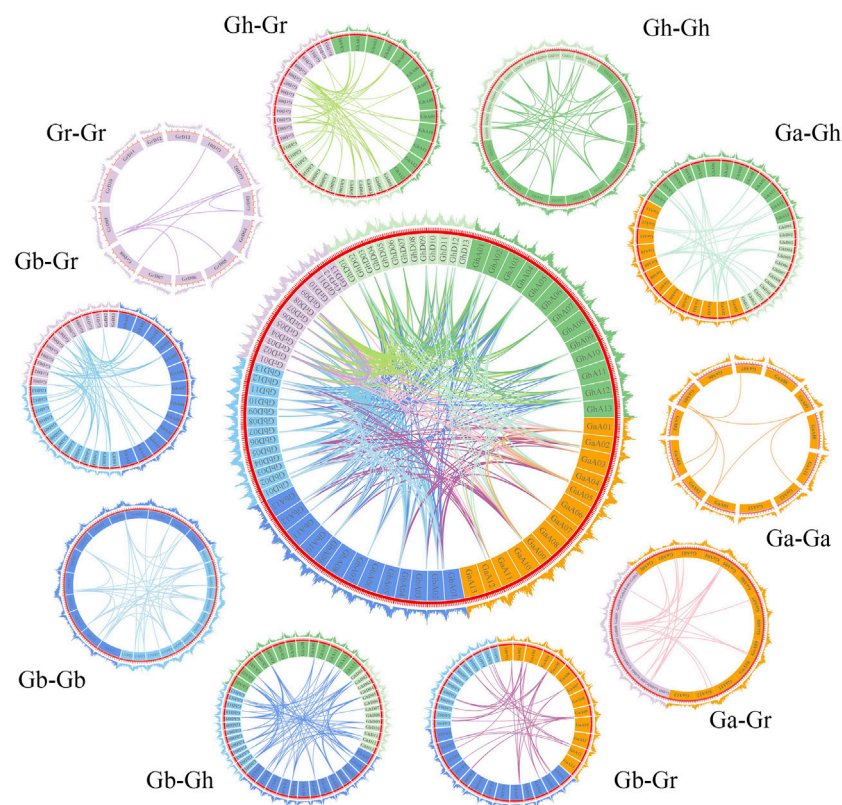


FIGURE 4

The collinearity relationship of repeated BSK gene pairs in the four cotton species of *Gossypium hirsutum*, *Gossypium barbadense*, *Gossypium arboreum* and *Gossypium raimondii*. The inner ring uses different colors to represent the chromosomes of different cotton species, the inner lines represent the collinearity of BSK genes, the outer ring represents the density of genes on the chromosome.

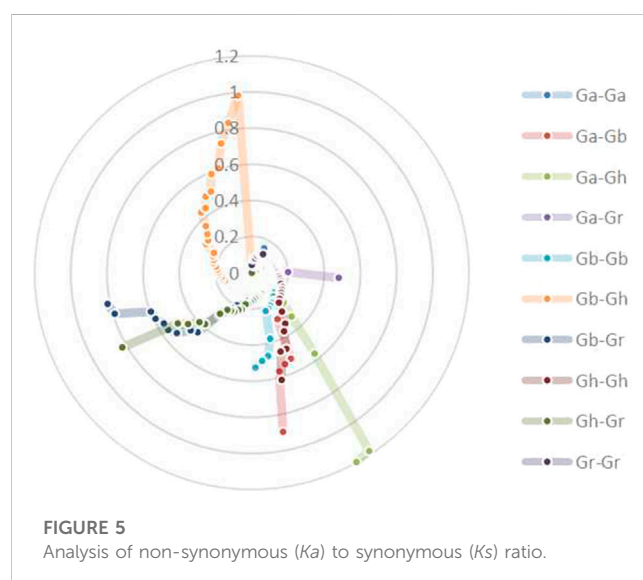


FIGURE 5

Analysis of non-synonymous (K_a) to synonymous (K_s) ratio.

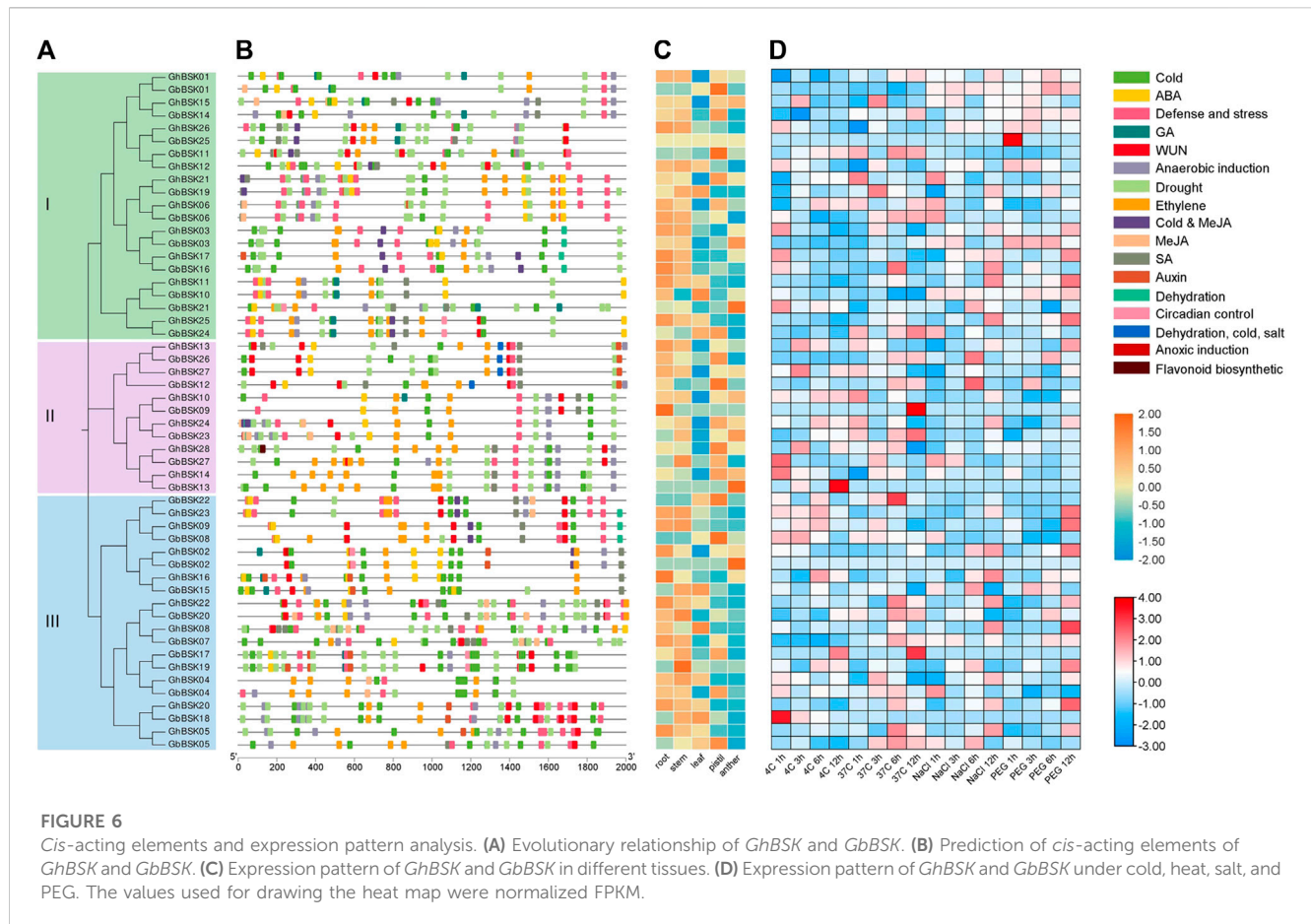
other three species. There is no gene distribution on the 10th, 12th, and 13th chromosomes. Only one BSK gene of *G. arboreum* is located on *Tig*, which means the evolution of BSK gene family is mature.

3.4 Conserved protein motifs, gene structure analysis and prediction of subcellular localization

Motif architecture, exon-intron structure analysis was performed on 83 BSKs of cotton, and their evolutionary relationships were visualized and drawn into a map (Figure 3). *GhBSK08*, *GhBSK18* and *GbBSK07* only lack PKC, and *GhBSK19* and *GhBSK22* only lack TPR. Both domains are missing from *GhBSK07*, *GbBSK21*, *GrBSK05*, and *GrBSK09*. The remaining 74 BSKs contain two domains. *GbBSK22* has two PKCs. The remaining 74 BSKs contain two domains, and their motifs are well conserved. Some BSK IIIs only retain a small number of motifs, which may be very critical to the function of BSK. Further, BSKs in the same clade were always predicted to have the same subcellular localization.

3.5 Gene duplication and collinearity analysis

To explore the evolutionary process of BSK genes, MCscanX was used to analyze the gene duplication patterns of four cotton species and conduct genetic correlation analysis (Figure 4). A



total of 385 gene pairs of genome-wide duplication were identified from 4 cotton species, of which 94 were segmental duplication and no tandem duplication was found. Based on this result, it is speculated that parallel homologous genes are generated through whole-genome duplication and segmental duplication and whole-genome duplication is an important driving force for species differentiation. Gh and Gb collinearity gene pairs are the most among the 10 groups, with 81 pairs, while Ga and Ga have the least, with only 7 pairs, which is in line with the comparison of the number of diploid and tetraploid genes. The other collinear pairs of Ga-Gb, Ga-Gh, Ga-Gr, Gb-Gb, Gb-Gr, Gh-Gh, Gh-Gr and Gr-Gr were 39, 35, 22, 41, 81, 59, 38, 55 and 8.

3.6 Selection pressure Ka/Ks is analyzed

To investigate the relationship between Darwinian positive selection and BSK gene duplication, the non-synonymous levels (Ka) and synonymous levels (Ks) were calculated for 385 duplicated gene pairs in 10 combinations from four cotton species. These combinations include Ga-Ga, Ga-Gb, Ga-Gh, Ga-Gr, Gb-Gb, Gb-Gh, Gb-Gr, Gh-Gh, Gh-Gr, Gr-Gr. Inference of selection pressure for duplicate gene pairs was based on Ka/Ks values. It is generally acknowledged that Ka/Ks = 1 indicates neutral selection (pseudogenes), Ka/Ks <

1 indicates purification or negative selection, which means that the gene evolution has a trend towards purification, and Ka/Ks > 1 indicates positive selection. Gene pairs with Ka/Ks > 1 appeared in all groups except Ga-Ga and Gr-Gr, indicating that these genes evolved faster. Most gene pairs have Ka/Ks values within 1, centered between 0 and 0.5 (Figure 5; Supplementary Table S2).

3.7 Cis-acting elements and expression pattern analysis

Cis-acting elements analysis of cotton showed that BSKs had a large number of hormone-responsive elements and stress-responsive elements induced by cold and drought, indicating that BSKs may be regulated by various hormones at different growth stages and participate in the response regulation of cotton to various stresses (Figure 6). Tissue-specific expression analysis showed that BSKs were highly expressed in roots and stems, and some BSK I and BSK II genes were highly expressed in pistil. BSK III had a high express level in leaf. Gene expression analysis under cold, heat, salt and PEG stresses showed that BSKs were differentially expressed under different stress treatments, which were associated with stress response elements.

AtBSK5 and ZmBSK1 acted as a positive regulator in plant salt tolerance (Li et al., 2012; Liu et al., 2022). In order to explore the

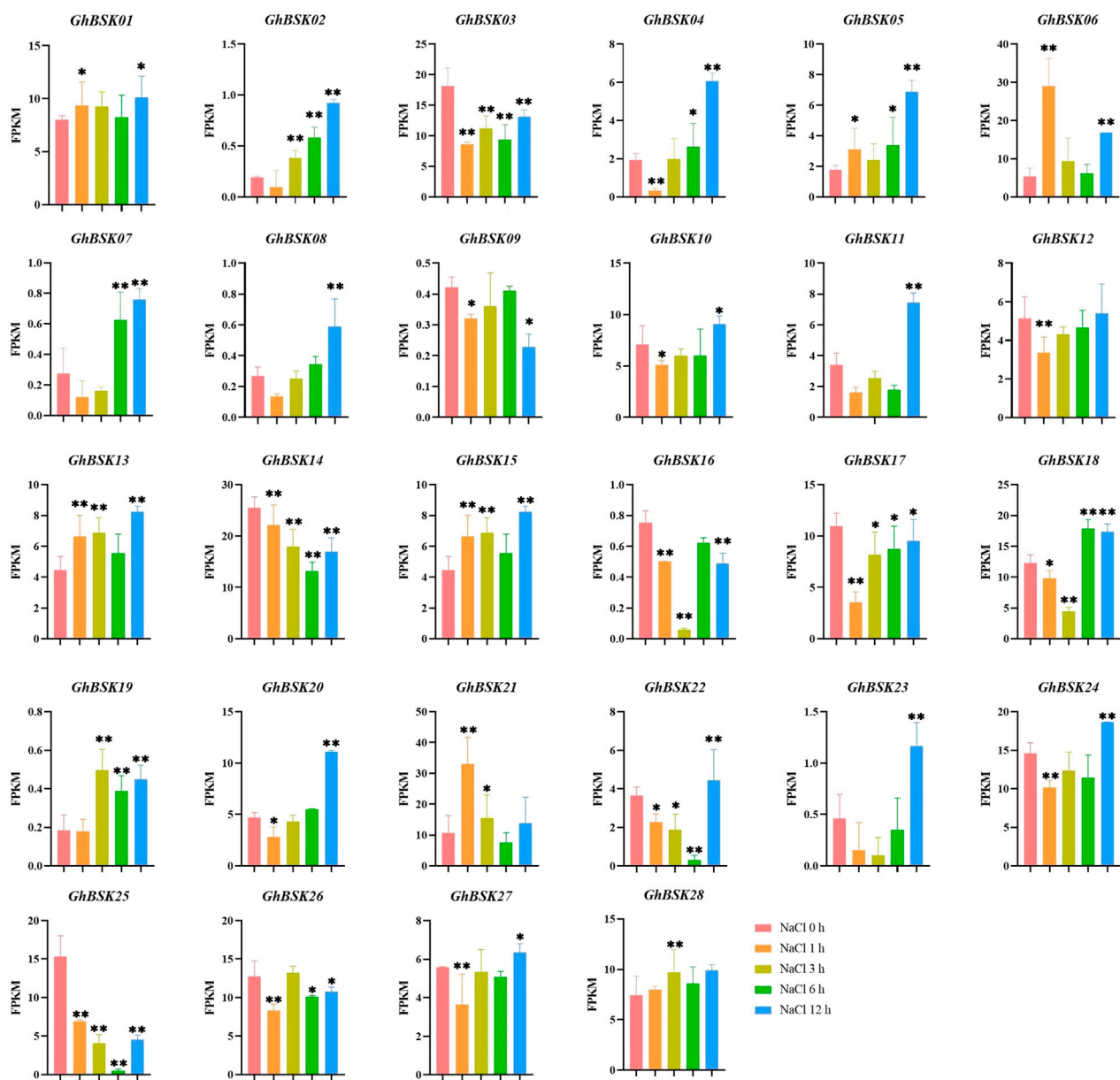


FIGURE 7
Expression analysis of *GhBSKs* under salt. The treatment groups were compared to NaCl 0 h group. Statistical significance was counted using one-way ANOVA. * mean $p < 0.05$, ** mean $p < 0.01$.

expression patterns of cotton BSKs under salt stress, the significance analysis of the expression levels of upland cotton BSK was carried out (Figure 7). The expression levels of *GhBSK03* and *GhBSK17* decreased significantly after salt stress and increased gradually with the increase of stress time. *GhBSK06* and *GhBSK21* increased at 1 h after salt stress and then decreased, while *GhBSK06* increased again at 12 h after treatment. *GhBSK24* decreased at 1 h after salt stress and increased at 12 h after treatment. *GhBSK10* did not change significantly at the initial treatment but increased significantly 12 h after treatment. Taken together, it implies that 12 h is an intentional node, and some BSKs cannot respond to salt stress in time.

4 Discussion

Several studies have reported the roles of BSK proteins from *Arabidopsis* and rice in BR signaling and immunity, as well as in abiotic stress responses (Mora-García et al., 2004; Tang et al., 2008; Bayer et al., 2009; Li et al., 2012; Shi et al., 2013; Sreeramulu et al., 2013; Wang et al., 2017; Yan et al., 2018; Yu et al., 2018; Ren et al., 2019). However, the knowledge of BSK proteins in cotton is still quite limited. In this study, a total of 152 BSKs were identified in four cotton species and other 11 plant species by a genome-wide search, and bioinformatics methods were used to analysis their evolution,

gene structure, promoter elements, protein structure and properties, gene expression patterns. Whole-genome duplication (WGD) has been reported in the two tetraploid cotton species studied (Paterson et al., 2012), and approximately two-fold amplification of the BSK gene was found. Genome duplication likely contributed the most to the expansion of the BSK gene family in many plant lineages. Approximately two-fold expansion of the BSK gene was found in plants with WGD, including *A. thaliana*, *P. trichocarpa*, and *Z. mays*. Furthermore, more than three-fold expansion of BSK was observed in *B. rapa*, possibly due to WTD (Li et al., 2019).

There are many similarities between cotton BSK genes. Except incomplete genes, BSKs possess an N-terminal PKC domain and a C-terminal TPR domain, consistent with previous studies (Tang et al., 2008). Although the expression patterns of BSKs were different, the homologous cotton BSKs proteins were basically identical in expression in tissues and organs, indicating that homologous proteins may have similar functions and redundancy.

The origin and development of the BR signaling system seems to be highly relevant with the evolution from aquatic to terrestrial plants, which have been observed in the ABA signaling system (Umezawa et al., 2010). A large number of ABA-responsive elements were also found in the promoter analysis of cotton BSKs (Figure 6). Besides ABA, a large number of hormone-responsive elements such as MeJA, GA, ethylene, SA and auxin were also found. It has been reported that BR interacts with ABA, Auxin, GA, JA and other hormones to regulate various growth and development processes of plants (Steber and McCourt, 2001; Assmann, 2004; Nemhauser et al., 2004; Perfus-Barbeoch et al., 2004; Hardtke et al., 2007; Finkelstein et al., 2008; Choudhary et al., 2012; Vandenbussche et al., 2013). BR is related to Auxin hypocotyl extension, root development, cell elongation, vascular bundle differentiation, lamellar tilt and shoot geotropism. Mutants of BR, which synergize with Auxin to promote cell elongation, often exhibit similar developmental defects, including an extremely short phenotype (Halliday, 2004; Teale et al., 2008). BR may affect plant stress response by promoting JA synthesis. BR also interacts with ethylene and auxin to regulate *Arabidopsis* shoot geotropism (Polko et al., 2013; Vandenbussche et al., 2013). Subcellular localization prediction of cotton BSKs revealed that no BSKs were anchored in the plasma membrane system.

Tissue-specific expression analysis of cotton BSKs suggests that different BSK proteins may function broadly in special tissues. The expression patterns of BSKs in the same subgroup were not always the same, and there were also differences in the expression patterns of *G. hirsutum* and *G. barbadense*. BR has been reported to play an important role in regulating root meristem maintenance, root elongation and promoting pollen germination and growth (González-García et al., 2011; Vogler et al., 2014). Significantly high expression of a large number of BSKs in the roots was observed in Figure 6. *GhBSK14*, *GbBSK01*, *GbBSK05*, *GbBSK08*, *GbBSK11*, *GbBSK12*, *GbBSK22*, and *GhBSK28* were only highly expressed in pistils, while *GbBSK02*, *GbBSK03*, *GbBSK13*, and *GbBSK21* were only highly expressed in anthers. Downstream genes of BSKs such as *Arabidopsis* BES1 and rice *OsBZR1* have been reported to be involved in anther and pollen development,

revealing that BSK members may be involved in male gametophyte development (Ye et al., 2010; Zhu et al., 2015). BRs also regulate photomorphogenesis through BES1 and associate with light signaling (Wang et al., 2012; Wang et al., 2018). A large number of light-responsive *cis*-acting elements were found in cotton BSK family members, suggesting that they may have light-dependent transcriptional regulation. Lots of BSKs were repressed by cold, heat, salinity and drought, only a few escaped, which may be related to their low expression in leaves. Some genes showed broad responses to multiple stresses, such as *GbBSK05* was highly expressed under heat and salt stress; *GbBSK11* and *GbBSK23* were highly expressed under cold and salt stress, and *GbBSK01* was highly expressed under salt and drought stress, and a few genes only responded to one abiotic stress. The upregulated expression of individual genes at different time points indicated that cotton BSKs played a regulatory role in cotton growth, development, and stress response. Further research is required to identify the interacting proteins of cotton BSK and investigate the molecular mechanisms for improving plant salt tolerance. Furthermore, the study will explore the role of cotton BSK in resisting other abiotic stresses to cultivate multi-resistance cotton.

5 Conclusion

In this study, the BSK family members of *G. hirsutum*, *G. barbadense*, *G. arboretum* and *G. raimondii* were identified for the first time, and their evolutionary, gene structure, conservative structure and expression pattern were analyzed. The evolution of the BSK family genes were very conservative and were divided into three clades, with different distribution in different species. In cotton, the extended family is mainly replicated by the whole genome, without tandem repeat events. The expression of BSK gene is regulated by various environments, such as light, plant hormones and stress. In conclusion, GhBSKs and GbBSKs may play an important regulatory role in cotton growth and development and stress response, but how to regulate hormone and stress response through BR signaling pathway needs further experiments to prove.

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

WY and YL conceived these experiments. YL performed these experiments and completed the manuscript with the assistance of all authors. YC, RC, XC, JuW, XuL, DW, SW, LG, YZ, CR, YF, MH, LZ, HZ, XiL, NX, JiW, HH, XF, YX, KN, MZ, and TJ provided critical feedback and revised manuscript. WY read and confirmed the final manuscript. All authors contributed to the article and approved the submitted version.

Funding

This research was supported by the Agricultural Science and Technology Innovation Program of Chinese Academy of Agricultural Sciences, China Agriculture Research System of MOF and MARA, and supported by National Natural Science Foundation of China (Grant No. 31901509).

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.

References

- Assmann, S. M. (2004). Plant G proteins, phytohormones, and plasticity: Three questions and a speculation. *Science's STKE* 2004 (264), re20. doi:10.1126/stke.2642004re20
- Bajguz, A., and Hayat, S. (2009). Effects of brassinosteroids on the plant responses to environmental stresses. *Plant physiology Biochem.* 47 (1), 1–8. doi:10.1016/j.plaphy.2008.10.002
- Bayer, M., Nawy, T., Giglione, C., Galli, M., Meinel, T., and Lukowitz, W. (2009). Paternal control of embryonic patterning in *Arabidopsis thaliana*. *Science* 323 (5920), 1485–1488. doi:10.1126/science.1167784
- Bent, A. F., and Mackey, D. (2007). Elicitors, effectors, and R genes: The new paradigm and a lifetime supply of questions. *Annu. Rev. Phytopathol.* 45, 399–436. doi:10.1146/annurev.phyto.45.062806.094427
- Boller, T., and Felix, G. (2009). A renaissance of elicitors: Perception of microbe-associated molecular patterns and danger signals by pattern-recognition receptors. *Ann. Rev. Plant Biol.* 60, 379–406. doi:10.1146/annurev.arplant.57.032905.105346
- Choudhary, S. P., Yu, J.-Q., Yamaguchi-Shinozaki, K., Shinozaki, K., and Tran, L.-S. P. (2012). Benefits of brassinosteroid crosstalk. *Trends plant Sci.* 17 (10), 594–605. doi:10.1016/j.tplants.2012.05.012
- Clouse, S. D., and Sasse, J. M. (1998). Brassinosteroids: Essential regulators of plant growth and development. *Annu. Rev. plant Biol.* 49 (1), 427–451. doi:10.1146/annurev.arplant.49.1.427
- Fàbregas, N., and Caño-Delgado, A. I. (2014). Turning on the microscope turret: A new view for the study of brassinosteroid signaling in plant development. *Physiol. Plant.* 151 (2), 172–183. doi:10.1111/ppl.12130
- Finkelstein, R., Reeves, W., Ariizumi, T., and Steber, C. (2008). Molecular aspects of seed dormancy. *Annu. Rev. Plant Biol.* 59, 387–415. doi:10.1146/annurev.arplant.59.032607.092740
- González-García, M.-P., Vilarrasa-Blasi, J., Zhiponova, M., Divol, F., Mora-García, S., Russinova, E., et al. (2011). Brassinosteroids control meristem size by promoting cell cycle progression in *Arabidopsis* roots. *Development* 138 (5), 849–859. doi:10.1242/dev.057331
- Grütter, C., Sreeramulu, S., Sessa, G., and Rauh, D. (2013). Structural characterization of the RLCK family member BSK8: A pseudokinase with an unprecedented architecture. *J. Mol. Biol.* 425 (22), 4455–4467. doi:10.1016/j.jmb.2013.07.034
- Halliday, K. J. (2004). Plant hormones: The interplay of brassinosteroids and auxin. *Curr. Biol.* 14 (23), R1008–R1010. doi:10.1016/j.cub.2004.11.025
- Hardtke, C. S., Dorcey, E., Osmont, K. S., and Sibout, R. (2007). Phytohormone collaboration: Zooming in on auxin–brassinosteroid interactions. *Trends Cell Biol.* 17 (10), 485–492. doi:10.1016/j.tcb.2007.08.003
- He, Z., Wang, Z.-Y., Li, J., Zhu, Q., Lamb, C., Ronald, P., et al. (2000). Perception of brassinosteroids by the extracellular domain of the receptor kinase BRI1. *Science* 288 (5475), 2360–2363. doi:10.1126/science.288.5475.2360
- Hu, Y., Chen, J., Fang, L., Zhang, Z., Ma, W., Niu, Y., et al. (2019). *Gossypium barbadense* and *Gossypium hirsutum* genomes provide insights into the origin and evolution of allotetraploid cotton. *Nat. Genet.* 51 (4), 739–748. doi:10.1038/s41588-019-0371-5
- Huang, H., He, Y., Cui, A., Sun, L., Han, M., Wang, J., et al. (2022). Genome-wide identification of GAD family genes suggests *GhGAD6* functionally respond to Cd stress in cotton. *Front. Genet.* 13, 965058. doi:10.3389/fgene.2022.965058
- Kim, T.-W., Guan, S., Burlingame, A. L., and Wang, Z.-Y. (2011). The CDG1 kinase mediates brassinosteroid signal transduction from BRI1 receptor kinase to BSU1 phosphatase and GSK3-like kinase BIN2. *Mol. Cell* 43 (4), 561–571. doi:10.1016/j.molcel.2011.05.037
- Kim, T.-W., Guan, S., Sun, Y., Deng, Z., Tang, W., Shang, J.-X., et al. (2009). Brassinosteroid signal transduction from cell-surface receptor kinases to nuclear transcription factors. *Nat. Cell Biol.* 11 (10), 1254–1260. doi:10.1038/ncb1970
- Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33 (7), 1870–1874. doi:10.1093/molbev/msw054
- Li, J., Li, Y., Chen, S., and An, L. (2010). Involvement of brassinosteroid signals in the floral-induction network of *Arabidopsis*. *J. Exp. Bot.* 61 (15), 4221–4230. doi:10.1093/jxb/erq241
- Li, Y., Zhang, H., Zhang, Y., Liu, Y., Li, Y., Tian, H., et al. (2022). Genome-wide identification and expression analysis reveals spinach brassinosteroid-signaling kinase (BSK) gene family functions in temperature stress response. *BMC genomics* 23 (1), 453. doi:10.1186/s12864-022-08684-5
- Li, Z.-Y., Xu, Z.-S., He, G.-Y., Yang, G.-X., Chen, M., Li, L.-C., et al. (2012). A mutation in *Arabidopsis* BSK5 encoding a brassinosteroid-signaling kinase protein affects responses to salinity and abscisic acid. *Biochem. biophysical Res. Commun.* 426 (4), 522–527. doi:10.1016/j.bbrc.2012.08.118
- Li, Z., Shen, J., and Liang, J. (2019). Genome-wide identification, expression profile, and alternative splicing analysis of the brassinosteroid-signaling kinase (BSK) family genes in *Arabidopsis*. *Int. J. Mol. Sci.* 20 (5), 1138. doi:10.3390/ijms20051138
- Liu, L., Sun, Y., Di, P., Cui, Y., Meng, Q., Wu, X., et al. (2022). Overexpression of a *Zea mays* brassinosteroid-signaling kinase gene ZmBSK1 confers salt stress tolerance in maize. *Front. plant Sci.* 13, 894710. doi:10.3389/fpls.2022.894710
- Mora-García, S., Vert, G., Yin, Y., Caño-Delgado, A., Cheong, H., and Chory, J. (2004). Nuclear protein phosphatases with Kelch-repeat domains modulate the response to brassinosteroids in *Arabidopsis*. *Genes and Dev.* 18 (4), 448–460. doi:10.1101/gad.1174204
- Nakaya, M., Tsukaya, H., Murakami, N., and Kato, M. (2002). Brassinosteroids control the proliferation of leaf cells of *Arabidopsis thaliana*. *Plant Cell Physiology* 43 (2), 239–244. doi:10.1093/pcp/pcf024
- Nemhauser, J. L., Mockler, T. C., Chory, J., and Dangel, J. (2004). Interdependency of brassinosteroid and auxin signaling in *Arabidopsis*. *PLoS Biol.* 2 (9), e258. doi:10.1371/journal.pbio.0020258
- Paterson, A. H., Wendel, J. F., Gundlach, H., Guo, H., Jenkins, J., Jin, D., et al. (2012). Repeated polyploidization of *Gossypium* genomes and the evolution of spinnable cotton fibres. *Nature* 492 (7429), 423–427. doi:10.1038/nature11798
- Perfus-Barbeoch, L., Jones, A. M., and Assmann, S. M. (2004). Plant heterotrimeric G protein function: Insights from *Arabidopsis* and rice mutants. *Curr. Opin. plant Biol.* 7 (6), 719–731. doi:10.1016/j.pbi.2004.09.013
- Polko, J. K., Pierik, R., van Zanten, M., Tarkowska, D., Strnad, M., Voesenek, L. A., et al. (2013). Ethylene promotes hyponastic growth through interaction with ROTUNDIFOLIA3/CYP90C1 in *Arabidopsis*. *J. Exp. Bot.* 64 (2), 613–624. doi:10.1093/jxb/ers356
- Ren, H., Willige, B. C., Jaillais, Y., Geng, S., Park, M. Y., Gray, W. M., et al. (2019). BRASSINOSTEROID-SIGNALING KINASE 3, a plasma membrane-associated scaffold protein involved in early brassinosteroid signaling. *PLoS Genet.* 15 (1), e1007904. doi:10.1371/journal.pgen.1007904
- Schwesinger, B., and Zipfel, C. (2008). News from the frontline: Recent insights into PAMP-triggered immunity in plants. *Curr. Opin. plant Biol.* 11 (4), 389–395. doi:10.1016/j.pbi.2008.06.001

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/fgene.2023.1169104/full#supplementary-material>

- Shi, H., Shen, Q., Qi, Y., Yan, H., Nie, H., Chen, Y., et al. (2013). BR-SIGNALING KINASE1 physically associates with FLAGELLIN SENSING2 and regulates plant innate immunity in *Arabidopsis*. *Plant Cell* 25 (3), 1143–1157. doi:10.1105/tpc.112.107904
- Sreeramulu, S., Mostizky, Y., Sunitha, S., Shani, E., Nahum, H., Salomon, D., et al. (2013). BSK s are partially redundant positive regulators of brassinosteroid signaling in *Arabidopsis*. *Plant J.* 74 (6), 905–919. doi:10.1111/tpj.12175
- Steber, C. M., and McCourt, P. (2001). A role for brassinosteroids in germination in *Arabidopsis*. *Plant physiol.* 125 (2), 763–769. doi:10.1104/pp.125.2.763
- Stout, M. J., Thaler, J. S., and Thomma, B. P. (2006). Plant-mediated interactions between pathogenic microorganisms and herbivorous arthropods. *Annu. Rev. Entomol.* 51, 663–689. doi:10.1146/annurev.ento.51.110104.151117
- Symons, G. M., Davies, C., Shavruk, Y., Dry, I. B., Reid, J. B., and Thomas, M. R. (2006). Grapes on steroids. Brassinosteroids are involved in grape berry ripening. *Plant physiol.* 140 (1), 150–158. doi:10.1104/pp.105.070706
- Tang, W., Kim, T.-W., Oses-Prieto, J. A., Sun, Y., Deng, Z., Zhu, S., et al. (2008). BSKs mediate signal transduction from the receptor kinase BRI1 in *Arabidopsis*. *Science* 321 (5888), 557–560. doi:10.1126/science.1156973
- Teale, W., Ditegou, F., Dovzhenko, A., Li, X., Molendijk, A., Ruperti, B., et al. (2008). Auxin as a model for the integration of hormonal signal processing and transduction. *Mol. Plant* 1 (2), 229–237. doi:10.1093/mp/ssn006
- Umezawa, T., Nakashima, K., Miyakawa, T., Kuromori, T., Tanokura, M., Shinozaki, K., et al. (2010). Molecular basis of the core regulatory network in ABA responses: Sensing, signaling and transport. *Plant Cell physiology* 51 (11), 1821–1839. doi:10.1093/pcp/pcq156
- Vandenbussche, F., Callebert, P., Zadnikova, P., Benkova, E., and Van Der Straeten, D. (2013). Brassinosteroid control of shoot gravitropism interacts with ethylene and depends on auxin signaling components. *Am. J. Bot.* 100 (1), 215–225. doi:10.3732/ajb.1200264
- Vogler, F., Schmalzl, C., Englhart, M., Bircheneder, M., and Sprunck, S. (2014). Brassinosteroids promote *Arabidopsis* pollen germination and growth. *Plant Reprod.* 27 (3), 153–167. doi:10.1007/s00497-014-0247-x
- Wang, J., Shi, H., Zhou, L., Peng, C., Liu, D., Zhou, X., et al. (2017). OsBSK1-2, an orthologous of AtBSK1, is involved in rice immunity. *Front. plant Sci.* 8, 908. doi:10.3389/fpls.2017.00908
- Wang, J., Zhang, Y., Xu, N., Zhang, H., Fan, Y., Rui, C., et al. (2021). Genome-wide identification of CK gene family suggests functional expression pattern against Cd²⁺ stress in *Gossypium hirsutum* L. *Int. J. Biol. Macromol.* 188, 272–282. doi:10.1016/j.ijbiomac.2021.07.190
- Wang, W., Lu, X., Li, L., Lian, H., Mao, Z., Xu, P., et al. (2018). Photoexcited CRYPTOCHROME1 interacts with dephosphorylated BES1 to regulate brassinosteroid signaling and photomorphogenesis in *Arabidopsis*. *Plant Cell* 30 (9), 1989–2005. doi:10.1105/tpc.17.00994
- Wang, Z.-Y., Bai, M.-Y., Oh, E., and Zhu, J.-Y. (2012). Brassinosteroid signaling network and regulation of photomorphogenesis. *Annu. Rev. Genet.* 46, 701–724. doi:10.1146/annurev-genet-102209-163450
- Yan, H., Zhao, Y., Shi, H., Li, J., Wang, Y., and Tang, D. (2018). BRASSINOSTEROID-SIGNALING KINASE1 phosphorylates MAPKKK5 to regulate immunity in *Arabidopsis*. *Plant Physiol.* 176 (4), 2991–3002. doi:10.1104/pp.17.01757
- Yang, H., Yang, Q., Zhang, D., Wang, J., Cao, T., Bozorov, T. A., et al. (2023). Transcriptome reveals the molecular mechanism of the ScALDH21 gene from the desert moss *Syntrichia caninervis* conferring resistance to salt stress in cotton. *Int. J. Mol. Sci.* 24 (6), 5822. doi:10.3390/ijms24065822
- Ye, Q., Zhu, W., Li, L., Zhang, S., Yin, Y., Ma, H., et al. (2010). Brassinosteroids control male fertility by regulating the expression of key genes involved in *Arabidopsis* anther and pollen development. *Proc. Natl. Acad. Sci. U. S. A.* 107 (13), 6100–6105. doi:10.1073/pnas.0912333107
- Yu, M.-H., Zhao, Z.-Z., and He, J.-X. (2018). Brassinosteroid signaling in plant–microbe interactions. *Int. J. Mol. Sci.* 19 (12), 4091. doi:10.3390/ijms19124091
- Zhang, H., Zhang, Y., Xu, N., Rui, C., Fan, Y., Wang, J., et al. (2021). Genome-wide expression analysis of phospholipase A1 (PLA1) gene family suggests phospholipase A1-32 gene responding to abiotic stresses in cotton. *Int. J. Biol. Macromol.* 192, 1058–1074. doi:10.1016/j.ijbiomac.2021.10.038
- Zhang, S., Hu, X., Dong, J., Du, M., Song, J., Xu, S., et al. (2022). Identification, evolution, and expression analysis of OsBSK gene family in *Oryza sativa Japonica*. *BMC plant Biol.* 22 (1), 565. doi:10.1186/s12870-022-03905-1
- Zhu, T., Liang, C., Meng, Z., Sun, G., Meng, Z., Guo, S., et al. (2017). CottonFGD: An integrated functional genomics database for cotton. *BMC Plant Biol.* 17 (1), 101. doi:10.1186/s12870-017-1039-x
- Zhu, X., Liang, W., Cui, X., Chen, M., Yin, C., Luo, Z., et al. (2015). Brassinosteroids promote development of rice pollen grains and seeds by triggering expression of Carbon Starved Anther, a MYB domain protein. *Plant J.* 82 (4), 570–581. doi:10.1111/tpj.12820



OPEN ACCESS

EDITED BY

Amaranatha Reddy Vennapusa,
Delaware State University, United States

REVIEWED BY

Prasad Parchuri,
Washington State University,
United States
Balpreet Kaur Dhatt,
Bayer Crop Science, United States
Karthik Putta,
ENKOCHEM, United States

*CORRESPONDENCE

Jinming He,
✉ hjm@squ.edu.cn
Yanhui Xiao,
✉ yhxiao@squ.edu.cn

[†]These authors have contributed equally
to this work

RECEIVED 17 May 2023

ACCEPTED 11 July 2023

PUBLISHED 27 July 2023

CITATION

Wang B, Wang Y, Yuan X, Jiang Y, Zhu Y,
Kang X, He J and Xiao Y (2023),
Comparative transcriptomic analysis
provides key genetic resources in clove
basil (*Ocimum gratissimum*) under
cadmium stress.
Front. Genet. 14:1224140.
doi: 10.3389/fgene.2023.1224140

COPYRIGHT

© 2023 Wang, Wang, Yuan, Jiang, Zhu,
Kang, He and Xiao. 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.

Comparative transcriptomic analysis provides key genetic resources in clove basil (*Ocimum gratissimum*) under cadmium stress

Bin Wang ^{1†}, Yukun Wang ^{1†}, Xiao Yuan¹, Yuanyuan Jiang¹,
Yunna Zhu¹, Xinmiao Kang^{1,2}, Jinming He^{1*} and Yanhui Xiao^{1*}

¹Guangdong Provincial Key Laboratory of Utilization and Conservation of Food and Medicinal Resources in Northern Region, Shaoguan Aromatic Plant Engineering Research Center, College of Biology and Agriculture, Shaoguan University, Shaoguan, China, ²College of Horticulture, South China Agricultural University, Guangzhou, China

Planting aromatic plant might be a promising strategy for safely utilizing heavy metal (HM)-contaminated soils, as HMs in essential oil could be completely excluded using some special technologies with ease. Clove basil (*Ocimum gratissimum* L.) is an important aromatic plant used in essential oil production. Improving cadmium (Cd) tolerance in clove basil can increase its production and improve the utilization efficiency of Cd-contaminated soils. However, the lack of genomic information on clove basil greatly restricts molecular studies and applications in phytoremediation. In this study, we demonstrated that high levels of Cd treatments (0.8, 1.6 and 6.5 mg/L) significantly impacted the growth and physiological attributes of clove basil. Cd contents in clove basil tissues increased with treatment concentrations. To identify Cd stress-responsive genes, we conducted a comparative transcriptomic analysis using seedlings cultured in the Hoagland's solution without Cd ion (control) or containing 1.6 mg/L CdCl₂ (a moderate concentration of Cd stress for clove basil seedlings). A total of 104.38 Gb clean data with high-quality were generated in clove basil under Cd stress through Illumina sequencing. More than 1,800 differential expressed genes (DEGs) were identified after Cd treatment. The reliability and reproducibility of the transcriptomic data were validated through qRT-PCR analysis and Sanger sequencing. KEGG classification analysis identified the "MAPK signaling pathway," "plant hormone signal transduction" and "plant-pathogen interaction" as the top three pathways. DEGs were divided into five clusters based on their expression patterns during Cd stress. The functional annotation of DEGs indicated that downregulated DEGs were mainly involved in the "photosynthesis system," whereas upregulated DEGs were significantly assigned to the "MAPK signaling pathway" and "plant-pathogen interaction pathway." Furthermore, we identified a total of 78 transcription factors (TFs), including members of bHLH, WRKY, AP2/ERF, and MYB family. The expression of six bHLH genes, one WRKY and one ERF genes were significantly induced by Cd stress, suggesting that these TFs might play essential roles in regulating Cd stress responses. Overall, our study provides key genetic resources and new insights into Cd adaption mechanisms in clove basil.

KEYWORDS

clove basil, cadmium stress, transcriptomic profiles, Cd-responsive genes, phytoremediation

Introduction

Heavy metals (HMs) are harmful environmental elements that are continuously released into the environment as a result of human and industrial activities (Jiang et al., 2022). The proliferation of HMs into the environment will cause soil and water contamination. Mining activity is one of most important reasons for HM pollution. For example, gold mining and processing are dominant industries in the Anka area of Nigeria. Such activities have caused a significant increase in HMs loading in soils and sediments in the local area (Adewumi and Laniyan, 2020). HMs-polluted soils lead to serious environmental issues and pose a threat to the safety of agricultural products (Zhu et al., 2018). Because most HMs could be absorbed by plants and stored in their sink tissues, eventually entering the food chain, and thereby posing a direct threat to human health (Caicedo-Rivas et al., 2022). To ensure the safety of agricultural products, the government typically imposes a moratorium on the use of HMs-polluted lands before the soil is completely remediated. This leads to a large amount of land being left wasted and idle, as it is reported that over 20% of farmlands are polluted by HMs in many countries, especially in developing countries (Xie et al., 2021). Therefore, the remediation or safe utilization of HMs-contaminated soils is important for limited farmland resources.

Remediation of HMs-contaminated soils mainly uses physical and chemical methods, such as immobilization, soil washing, chemical leaching, and stabilization (Faizan et al., 2022). These methods aim to remove HMs from the soil, or to reduce HMs concentration to a safe level, or to inactivate HM mobility. Nevertheless, these methods are generally expensive or detrimental to soil properties in practice (Zhu et al., 2018). Therefore, alternative and sustainable methods need to be developed. Phytoremediation is a biological process that uses plants to remove or enrich HMs from the soil (Simmer and Schnoor, 2022). However, the use of most hyper-accumulating plants, such as *Sedum alfredii* Hance, tall fescue and some plant species in the *Brassicaceae* and *Crassulaceae* families, has been limited due to their lack of economic value and small biomass as well as the long time required for remediation (Roosens et al., 2008; Gao et al., 2013; Dong et al., 2019). One potential strategy to address this issue is the remediation and simultaneous utilization of HMs-contaminated soils using some cash plants that can accumulate HMs in their tissues without suffering from toxicity.

Cultivating non-edible plants with high economic values for industrial purposes may be a feasible strategy to both safely utilize and remediate HMs-contaminated soils (Wang et al., 2023). Planting industrial plants can reduce the risk of human exposure to HMs through the food chain. Moreover, these plants can also have remediation effects on HMs-polluted soils by absorbing and accumulating HMs from the soil. Aromatic plants are cash crops used for essential oil extraction (Saeed et al., 2014), an important material in various product productions (Ekiert et al., 2021). Furthermore, HMs in essential oil can be readily removed through several special methods such as cold pressing and steam

distillation (Gautam and Agrawal, 2017). Studies have reported that several aromatic plant species, such as peppermint, vetiver, and lemon grass, are highly tolerant to HMs, and HM stress could increase essential oil percentage and the synthesis of other metabolites, such as phenolic compounds (Pandey et al., 2019; do Prado et al., 2022; Wang et al., 2023). These reports suggest that planting aromatic plants is a sustainable and suitable method to utilize HMs-contaminated soils due to the monetary benefits.

Cadmium (Cd) is one of the most dangerous pollutants among HMs in nature and is widespread in soils (Zhang et al., 2023). As a toxic metallic element, Cd ions have been shown to negatively impact plant growth, even at very low concentrations (Jiang et al., 2020), resulting in reduced biomass and essential oil yield. Therefore, improving the Cd tolerance of aromatic plants through genetic engineering is crucial for the effective utilization of Cd-contaminated soils. Clove basil (*Ocimum gratissimum*) is an important species in the *Lamiaceae* family, because it produces essential oil in shoots and leaves (Ugbogu et al., 2021). Clove basil is a perennial shrub distributed in tropical and subtropical regions worldwide and can grow up to 1–3 m in height (Parasar et al., 2021). Due to its strong adaptability and resilience, it can grow in various soil types (Vilanova et al., 2018). However, to our best knowledge, available studies on how Cd stress influences clove basil performance are rare (do Prado et al., 2022). Furthermore, genomic data from *Ocimum* species are lacking, and available transcriptomic data is very limited (Parasar et al., 2021), with fewer than 150 ESTs deposited in the NCBI database. The lack of sufficient genomic data significantly hinders molecular studies in this plant species.

For plant species without sequenced genomes, RNA-seq is a relatively low-cost, simple, and efficient method to gather functional genomic data (Agarwal et al., 2014). This technique has been widely used to investigate transcriptomic responses to abiotic stress in several aromatic plants, such as peppermint (Wang et al., 2023) and *Artemisia annua* (Vashisth et al., 2018). We aim to investigate the effects of Cd treatment on the growth performance of clove basil seedlings and to present the first transcriptome profiles of clove basil leaves under Cd stress conditions at a global level using Illumina sequencing technology combined with bioinformatics analysis. Thus, the goals of this study are: 1) to investigate the influences of Cd stress treatments on growth and physiological attributes of clove basil; 2) to conduct transcriptomic profiles to a genome-wide extent; 3) to identify key genetic resources from clove basil under Cd stress for future molecular studies. Additionally, the functional analysis of Cd-responsive genes could provide valuable insights into the adaption mechanisms under Cd stress in plants.

Materials and methods

Materials, seedling cultivation and Cd stress treatment

The experimental setup was shown in Supplementary Figure S1. Clove basil (*Ocimum gratissimum*) seeds were sown in seedling-raising plates of matrix mixed with peat and perlite in a proportion

1:1. The seedlings, which had been grown for 14 days after germination and reached a height of approximately 8 cm, were individually transplanted into seedling-raising plates filled with a mixed matrix, ensuring ample space for further growth (each hole contained one seedling). After 21 days of growth in these plates, the seedlings were then transplanted into cuboid plastic containers filled with Hoagland's nutrient solution. Prior to the Cd treatment, the seedlings were acclimatized in the nutrient solution for a period of 1 week (7 days). Each container contained 20 lines of seedlings and regarded as one biological replicate. All seedlings were grown in a greenhouse at Shaoguan university, Shaoguan, China, under a 16/8 h light/dark photoperiod and 23°C.

Different dosages of CdCl₂ (Aladdin, China) solutions were separately added to the Hoagland's nutrient solution (Supplementary Figure S1). The final concentrations of Cd ions in each treatment were as follows: 0 (control), 0.4, 0.8, 1.6, and 6.5 mg/L for Cd treatments, respectively. In a previous study, we investigated the effects of different intensities of Cd treatments on peppermint growth (Wang et al., 2023), as both peppermint and clove basil belong to the *Labiatae* family. Therefore, in this study, we designed the Cd concentration of stress treatments based on our previous study to test Cd responses in clove basil seedlings. Each treatment group contained 60 seedling lines (three biological replicates).

During Cd treatment, leaf, shoot and root samples of clove basil seedlings were collected separately at 0 h, 24 h (1 day), 72 h (3 days) and 14 d (2 weeks) for further measurements or investigations. For sample collection, the samples from nine lines of one treatment were combined into a sample. After sampling, samples were quickly frozen within liquid nitrogen and stored at -80°C.

Determination of Cd contents in clove basil seedlings

The collected clove basil samples (leaf, shoot, and root) were dried in an air-circulating oven at 105°C for 10 min and at 80°C until a constant weight was reached. Then, the samples were ground into powder and sieved through a 100 mesh sieve. 200 mg of powders were digested with 10 mL of high-purity nitric acid-perchloric acid solutions (4:1, v/v). Finally, the digested solutions were added up to 100 mL using 2% nitric acid. Cd contents in clove basil samples were measured using a flame atomic absorption spectrometer (A3-AFG, PERSEE, China). The samples were grouped into roots, shoots, and leaves for determining Cd concentrations, respectively. The result of Cd content was expressed as milligram per kilogram dry weight (mg kg⁻¹ DW) (Gao et al., 2013).

Evaluation of the effects of Cd treatments on the growth of clove basil seedlings

Phenotypic changes of clove basil under different intensities of Cd treatments were analyzed first to test their Cd tolerance before RNA-Seq. The effects of different concentrations of Cd treatments on the seedling growth of clove basil were evaluated at 14 days after Cd treatment. The height of the seedlings was measured using a scale ruler, and the seedlings were dried to a constant weight and then used for weight measurements.

Chlorophyll relative contents (SPAD) in leaves under different levels of Cd stress were measured using a portable plant nutrient analyzer (TYS-4N, China). The contents of chlorophyll pigments in leaves of clove basil were determined using the method developed by Lichtenthaler (1987) through a UV spectrophotometer (UV-6100, SHjingmi, China). The root system of clove basil seedling was pictured and analyzed with a MICROTEK scanner (MRS-9600 TFU2L, China) (Wang et al., 2023).

Analysis of physiological parameters of clove basil under Cd stress

The activities of superoxide dismutase (SOD), catalase (CAT), and peroxidase (POD) in clove basil leaves were measured using a previously described method (Wang et al., 2023; Yuan et al., 2023). Total soluble protein (TSP) content in clove basil leaves was determined with the Coomassie brilliant blue method (Snyder and Desborough, 1978).

Total RNA extraction

The total RNA from clove basil leaves was isolated using an RNAprep Pure Plant Plus Kit (Tiangen, China) (Xiao et al., 2021). RNA quantification and qualification were performed as previously described (Wang et al., 2022).

Unlike other Cd-accumulating plants that lack economic value, clove basil is commonly used for extracting essential oil, which is primarily produced and stored in the leaves. Furthermore, the transport of Cd ions to the cytoplasm and vacuole of plant leaves through HM transporters is one of significant mechanisms for Cd detoxification. One of our objectives is to explore potential mechanisms of Cd detoxification in clove basil, a commercially valuable plant that produces essential oil, which might differ from those in other plants under Cd stress. Therefore, we focused our investigations on the transcriptome responses of clove basil leaves in response to Cd stress in this study.

cDNA library construction and RNA sequencing

The preparation of cDNA library was carried out using a NEB Next Ultra RNA Library Prep Kit for Illumina (NEB, United States) following the manufacturer's recommendations at the Biomarker Biotechnology Corporation (Beijing, China). The detailed instructions for library construction were described in a previous study (Wang et al., 2021a). The quality of all cDNA libraries was assessed using a Bioanalyzer 2100 system (Agilent Technologies, United States). Before transcriptome sequencing, a cBot Cluster Generation System with TruSeq PE Cluster Kit v4-cBot-HS (Illumina, United States) was employed to generate clusters. Finally, all constructed libraries were sequenced using the Illumina HiSeq2500 system, and paired-end reads were generated and converted to fastq format using bcl2fastq tool (Illumina, United States).

Transcriptomic data analysis and the identification of differential expressed genes (DEGs)

The raw reads were processed first through in-house perl scripts to obtain clean reads by removing those containing adapters and poly-N. At this step, GC-contents, Q20, Q30, and sequence duplication level in the clean data were calculated. We used Trinity software for transcriptome assembly (Zhu et al., 2018).

Expectation maximization tool was used to quantify the expression values of assembled transcripts, and we calculated fragments per kilobase of transcript per million fragments mapped (FPKM). To compare the differences in expression levels between two tested groups (each consisting of three replicates), we used the DESeq2 package. Differential expressed genes (DEGs) were identified based on the following thresholds: Q-value (adjusted p -value) ≤ 0.01 and fold change (FC) ≥ 1.5 (Wang et al., 2023). Each treatment had three biological replicates. Transcriptomic data was compared at specific time points to screen for DEGs: treatment at 24 h versus control at 24 h, and treatment at 72 h versus control at 72 h. FPKM values were utilized to demonstrate the expression patterns of DEGs during Cd treatment at different time intervals (0, 24, and 72 h).

Functional annotation and classification of the detected DEGs

The assembled sequences were searched against the NR protein database in the NCBI (<ftp://ftp.ncbi.nlm.nih.gov/blast/db/>) using BLASTX algorithm. Other public databases including Gene Ontology (GO) (<http://www.geneontology.org/>), Kyoto Encyclopedia of Genes and Genomes (KEGG) (<http://www.genome.jp/kegg/>), Swiss-Prot (<http://www.uniprot.org/>), Pfam (<https://www.ebi.ac.uk/interpro/>), and KOG/COG (<http://www.ncbi.nlm.nih.gov/COG/>) were also employed to annotate gene function or classification.

The quality validation of transcriptomic data by qRT-PCR analysis and Sanger sequencing

To validate the expression patterns of DEGs during Cd stress, we employed qRT-PCR analysis, and nine representative DEGs involved in Cd stress response were randomly selected. qRT-PCR analysis was conducted using previous methods (Xiao et al., 2022). The relative expression level of a specific gene was calculated with the $2^{-\Delta\Delta CT}$ method developed by Schmittgen and Livak (2008). An annotated gene known as *Actin 7* (c52716.graph_c0) was used as an internal control gene to perform normalization in this study. The expression levels of this *Actin* gene were not significantly influenced by Cd treatment (Supplementary Figure S2), indicating that it could be used as an internal control gene. The results of the expression validation of DEGs also demonstrated the reliability and reproducibility of this *Actin* gene as an internal reference gene (Figure 4).

Five representative DEGs were selected for PCR amplification and Sanger sequencing. First, the full-length sequence of each gene was cloned into pUCm-T vectors. Subsequently, the recombinant

vectors were transformed into *E. coli* (DH-5 α). The positive clones were validated again by PCR amplification, and then sequenced through Sanger sequencing. The sequences were analyzed using the BLAST program for identity alignment of assembled transcripts with Sanger sequences.

Specific primers used in this study were designed with an online tool, Primer-BLAST (<https://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi>), and listed in Supplementary Table S1.

Statistical analysis

The data in this study were expressed as the mean \pm standard deviation (SD), and values were calculated with the Excel 2016 software (Microsoft, United States). To analyze statistical differences between different samples, we used SPSS 22.0 software (IBM, United States) and employed a Student's t -test with $p \leq 0.05$ considered statistically significant. The statistical differences among samples were indicated with lowercase letters above bars.

Results

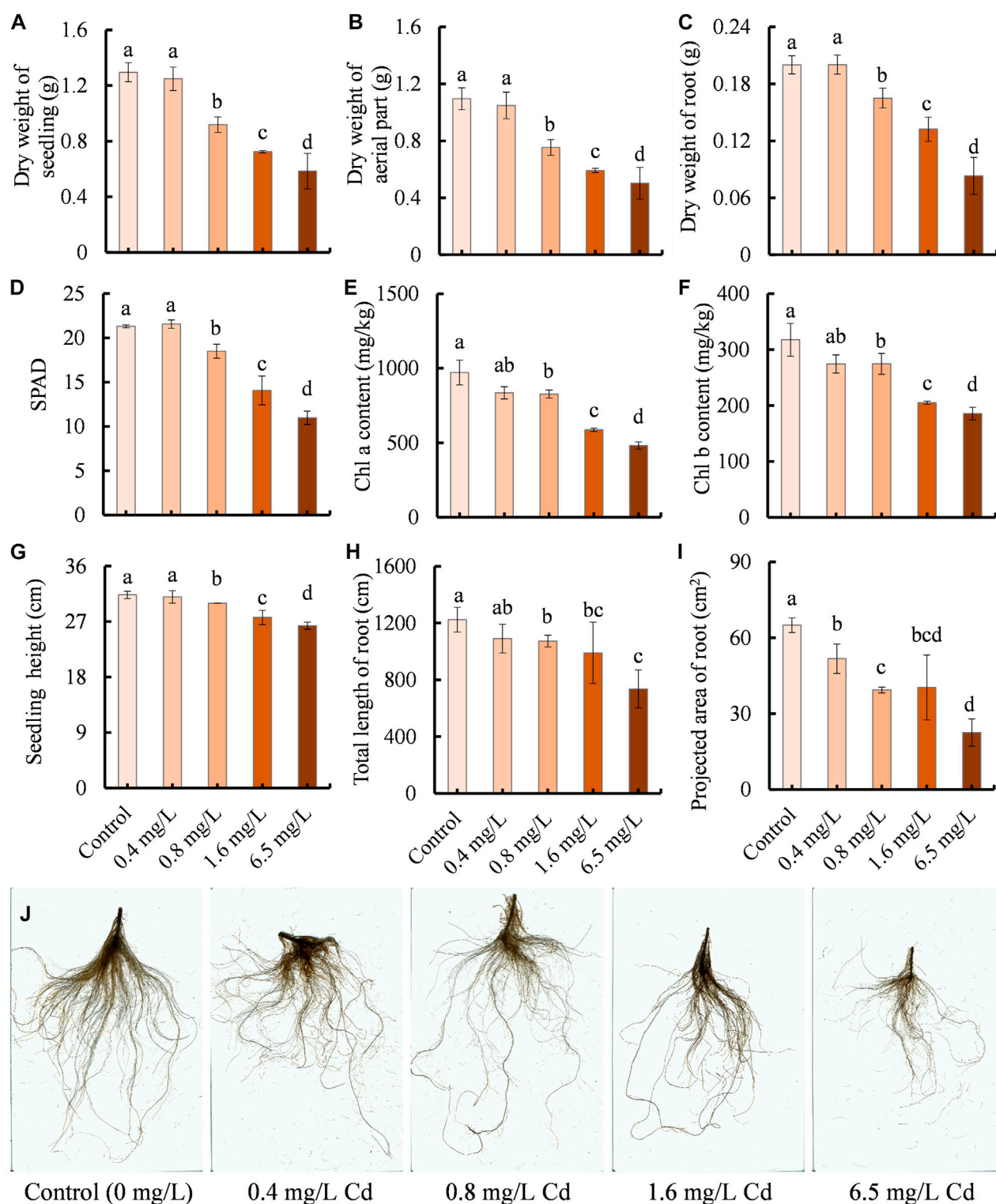
Effects of Cd treatment on the growth of clove basil

The growth of clove basil was evaluated under different concentrations of Cd treatments by measuring dry weight, chlorophyll contents, seedling height, and root development. Cd treatments led to significant reductions in dry weight, chlorophyll contents, height, and root system development of clove basil (Figure 1). Compared with the control group, seedlings exposed to Cd concentrations exceeding 0.4 mg/L had decreased values ranging from 4.26% to 65.31% (Figure 1). High levels of Cd treatments caused severe Cd injuries on young leaves (Supplementary Figure S3).

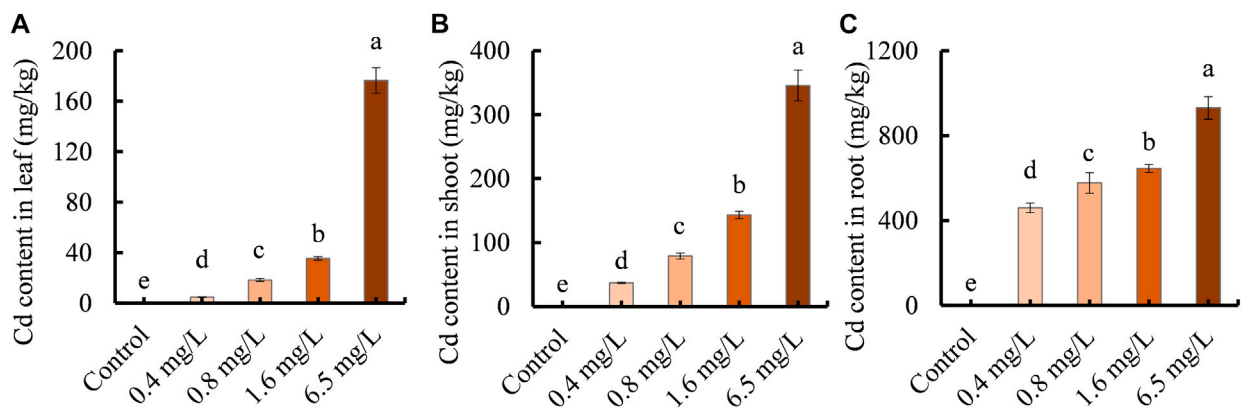
No significant differences were observed between the control and 0.4 mg/L Cd treatment in terms of dry weight, chlorophyll contents, seedling height, and root length (Figures 1A–I). Furthermore, no visible differences in the root morphology were found between the control and 0.4 mg/L Cd treatment and no Cd injuries on young leaves were found in the control and 0.4 mg/L Cd treatment (Figure 1J; Supplementary Figure S3). These results suggest that clove basil has the potential to withstand mild Cd stress, making it useful for Cd-contaminated soil utilization.

The contents of Cd ions in clove basil tissues

We investigated the ability of clove basil seedlings to accumulate or enrich Cd ions by measuring Cd ion contents in leaves, shoots, and roots following treatment with different concentrations of CdCl₂. No Cd ions were detected in the control group tissues. Cd ion contents in all tissues increased with increasing Cd treatment concentration, with the highest Cd accumulation observed in leaf, shoot, and root tissues at 6.5 mg/L Cd treatment (Figure 2). Notably, root tissues had significantly higher Cd contents than leaf and shoot tissues under various Cd treatments (Figure 2), suggesting that clove basil accumulated much less Cd ions in the above ground part.

**FIGURE 1**

Effects of Cd treatments with different concentrations (0, 0.4, 0.8, 1.6, and 6.5 mg/L) on the growth performance of clove basil. (A), dry weight of seedlings (whole plant); (B), dry weight of aerial part; (C), dry weight of root; (D), relative contents of chlorophyll (SPAD); (E,F), contents of chlorophyll a and b pigments, respectively; (G), seedling height; (H), total length of root; (I), projected area of root; (J), root morphology. The growth performance of clove basil seedlings was evaluated following 2 weeks (14 days) of Cd treatments. Each value is presented as means \pm standard error from three repeats ($n = 3$). Statistical differences ($p \leq 0.05$) between treatments were compared using the SPSS software and indicated using different letters above the bars.

**FIGURE 2**

Cd ion contents in clove basil tissues following 2 weeks of Cd treatments (0, 0.4, 0.8, 1.6, and 6.5 mg/L). (A), Cd ion contents in leaves; (B), Cd ion contents in shoots; (C), Cd ion contents in roots. Each value is presented as means \pm standard error from three repeats ($n = 3$). Statistical differences ($p \leq 0.05$) between treatments were compared using the SPSS software and indicated using different letters above the bars.

RNA-seq data analysis and the identification of DEGs under Cd stress

In order to investigate physiological and molecular changes of clove basil under Cd stress, a moderate stress concentration (1.6 mg/L) was selected to treat clove basil seedlings for RNA-seq analysis. A total of 15 cDNA libraries including 5 treatment groups (control at 0, 24 and 72 h, Cd treatment at 24 and 72 h, three biological replicates) were constructed. The average number of pair-end raw reads among the 15 samples was 22,147,978 (Supplementary Table S2). Quality evaluation of 15 samples generated 104.38 Gb data and the mean size of 15 cDNA libraries was 6.96 Gb. After filtration, GC contents of all samples varied from 48.72% to 50.81%, and Q30 values varied from 92.64% to 93.88% (Supplementary Table S2). The length of 76.88% transcripts was over 500 bp, and more than 58.00% transcripts were over 1 kb (Supplementary Table S3). Transcriptome assembly generated 79,637 transcripts (Supplementary Data Sheet S1) and 51.56% (41,060) unigenes were successfully annotated against nine public databases (Supplementary Table S4). Mapping efficiency of 15 samples varied from 75.34% to 77.27% (Supplementary Table S5). The results indicated that these high-quality transcriptome data can be further used for bioinformatics analysis.

The expression differences of specific gene were compared to the control at each sampling point (24 and 72 h, respectively) (Figure 3; Supplementary Data Sheets S2, S3). The DEGs were divided into up or downregulated transcripts based on regulation manner. Totally, 1,115 (605 up and 510 downregulated) DEGs were identified in Cd-treated leaves compared with control groups at 24 h time point (Figure 3A), while 1,156 (537 up- and 619 downregulated) were found at 72 h (Figure 3B). Cd treatment induced more DEGs at 24 h of initial stage, whereas more DEGs were repressed at 72 h (Figure 3C).

At 24 h, Cd treatment specifically affected the expressions of 663 DEGs (395 up and 268 downregulated), while at 72 h, 704 DEGs (327 upregulated and 377 downregulated) were specifically affected (Figure 3C). Additionally, 452 overlapped genes (210 up and 242 downregulated) were differentially expressed at both time

points (Figure 3C). The overall expression profile of these 452 DEGs was shown in Supplementary Figure S4, with most showing remarkable differences between Cd treatment and control.

The quality evaluation of transcriptomic data

qRT-PCR analysis, PCR amplification and Sanger sequencing of random genes were carried out to verify the reproducibility and credibility of RNA-seq data. Four DEGs (*OT3*, *ZAT10*, *HSP90*, and *HSP20*) showed steady increase, while three DEGs (*AAT*, *WAT1*, and *GST3*) exhibited constant reduction during Cd treatment in RNA-seq data. The expressions of two DEGs, *GST3*, and *SOD[Cu-Zn]*, were only induced by Cd treatment at 72 h (Figure 4). qRT-PCR validation of these nine representative DEGs showed high similarity with RNA-Seq data (Figure 4), indicating that the transcriptomic data could really, instantly reflect transcriptomic changes under Cd stress.

The ORF sequences of five DEGs were successfully amplified through PCR and sub-cloned to T vectors for Sanger sequencing, which matched closely to RNA-seq data with only a few bases not exactly matching (Supplementary Figure S5). This provides further evidence for the accuracy and reliability of the transcriptome sequencing data.

Functional annotation of Cd-responsive genes with COG and GO database

Phylogenetic classification was performed using Orthologous Groups (COG) analysis. A total of 392 DEGs generated by Cd treatment for 24 h were matched and grouped into 20 functional classes, with “signal transduction mechanisms” (43, 10.97%) and “carbohydrate transport and metabolism” (37, 9.44%) being the largest groups in percentage (Supplementary Figure S6A; Supplementary Data Sheet S2). Similarly, 460 DEGs from Cd treatment for 72 h were matched and classified into 22 functional classes, with “general function prediction only” (49, 10.65%),

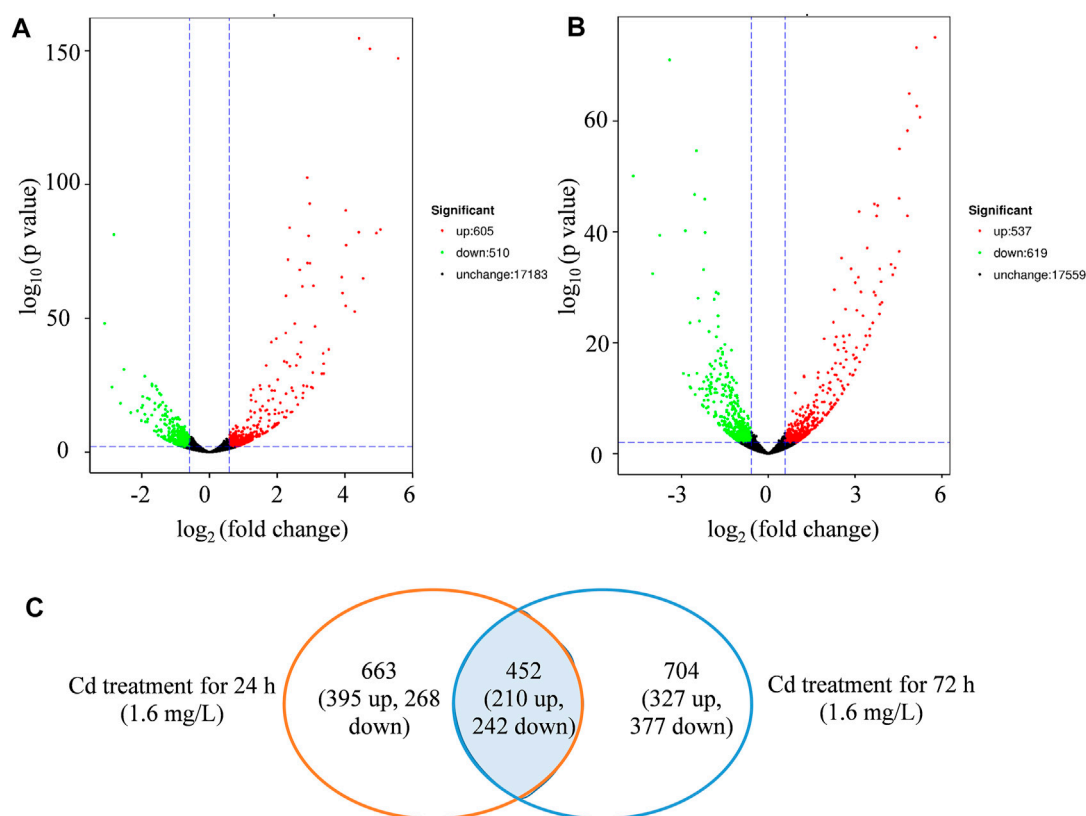


FIGURE 3

The identification of differentially expressed genes (DEGs). (A), volcano map of DEGs regulated by Cd treatment (1.6 mg/L) at 24 h. (B), volcano map of DEGs regulated by Cd treatment (1.6 mg/L) at 72 h. (C), venn diagram of DEGs regulated by treatment (1.6 mg/L) in the comparison between 24 and 72 h.

“carbohydrate transport and metabolism” (44, 9.57%), and “posttranslational modification, protein turnover, chaperones” (42, 9.13%) being the three largest groups in percentage (Supplementary Figure S6B; Supplementary Data Sheet S3).

The GO database were also used to classify the potential functions of annotated DEGs. Cd-treatment-caused DEGs were assigned to 45 functional groups under three major categories, showing similar classifications at both 24 h and 72 h (Supplementary Figure S7). In the biological process category, “metabolic process” and “cellular process” had most DEGs assigned. In the cellular component category, the terms of “cell,” “cell part,” “membrane,” “membrane part and organelle” were dominant terms, while “binding” and “catalytic activity” were the most common clusters in the molecular function category (Supplementary Figure S7).

The KEGG pathway analysis and expression patterns of Cd-responsive genes in clove basil

KEGG pathway analysis was further performed to predict the functions of Cd-responsive genes. A total of 519 DEGs were assigned to 50 specific KEGG pathways at 24 h, while 555 DEGs were assigned to 53 KEGG pathways at 72 h (Figure 5A). At 24 h, the

“plant-pathogen interaction pathway” had the most DEGs assigned (68, 13.10%), followed by the “MAPK signaling pathway” (40, 7.71%) and the “plant hormone signal transduction” pathway (32, 6.17%). At 72 h, most DEGs were classified into pathways of “carbon metabolism” (31, 5.59%) and “photosynthesis” (29, 5.23%) (Figure 5A). The results implied that long term of Cd exposure damaged photosynthetic system, which led to insufficient energy supply for survival and ultimately resulted in cell death under Cd stress.

The Cd-responsive genes were classified into five groups based on their expression patterns during Cd treatment, with 538 DEGs in cluster 1, 669 DEGs in cluster 2, 1,284 DEGs in cluster 3, 1,121 DEGs in cluster 4, and 321 DEGs in cluster 5 (Figure 5B). As Cd-induced impacts on clove basil were positive or negative, we focused on analyzing the functions of DEGs that were constantly induced (cluster 2 DEGs) or repressed (cluster 1 DEGs) through KEGG enrichment analysis.

DEGs in the cluster 1 showed significant downregulation in expression due to Cd stress and were enriched in three KEGG pathways: “porphyrin and chlorophyll metabolism” (17 genes), “photosynthesis” (24 genes) and “photosynthesis-antenna protein” (18 genes) (Figure 5C). This conforms that Cd stress significantly reduced photosynthetic efficiency in clove basil. On the other hand, upregulated DEGs (cluster 2) were significantly enriched in “plant-pathogen interaction” pathway (13 genes) and

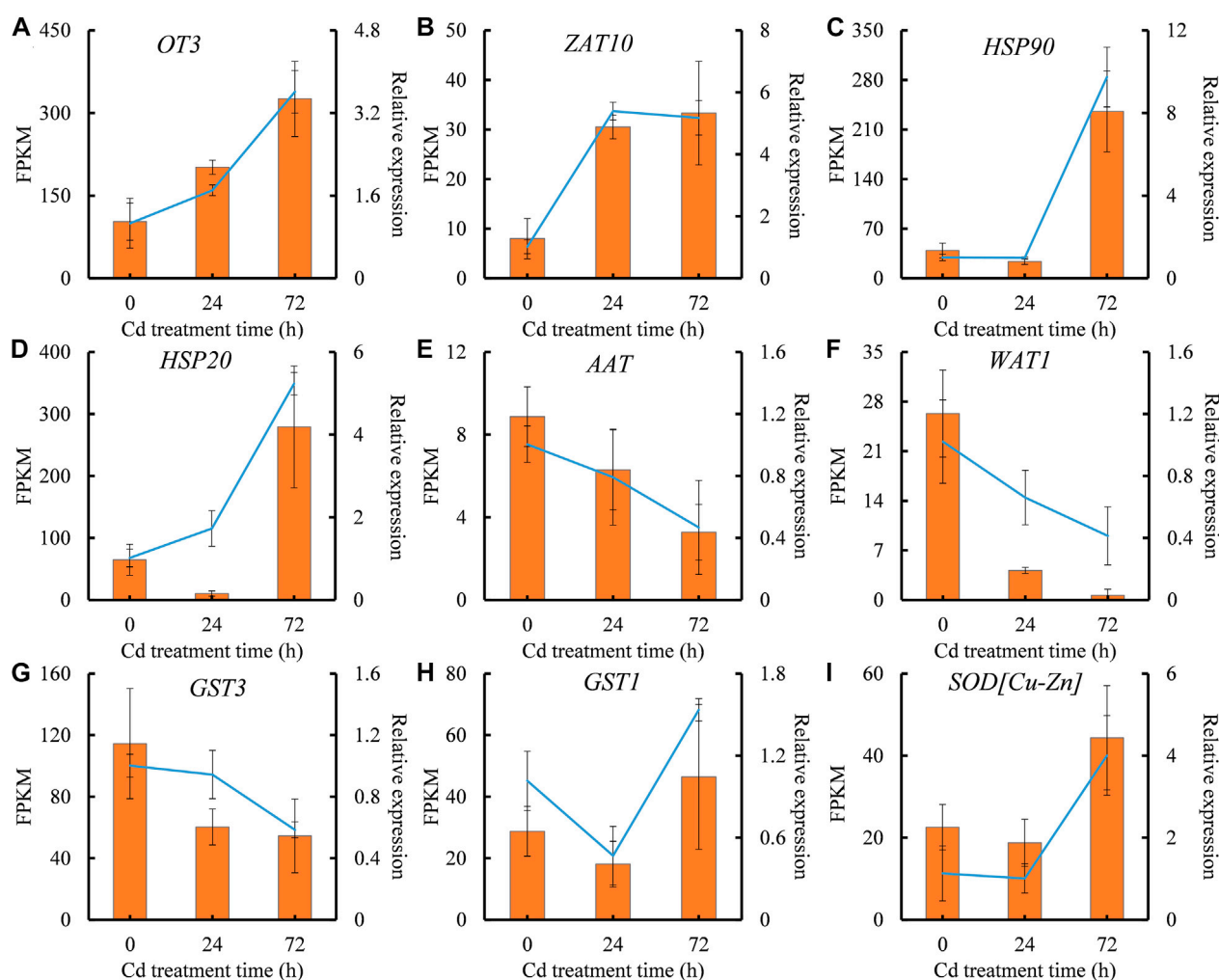


FIGURE 4

The expression patterns of representative differentially expressed genes (DEGs) validated by qRT-PCR analysis. (A), *OT3* expression. (B), *ZAT10* expression. (C), *HSP90* expression. (D), *HSP20* expression. (E), *AAT* expression. (F), *WAT1* expression. (G), *GST3* expression. (H), *GST1* expression. (I), *SOD [Cu-Zn]* expression. The X-axis represents treatment time. The Y-axis on the left represents the expression level of specific gene obtained from RNA-seq (histogram), and on the right, from qRT-PCR analysis (line chart). Each value is presented as means \pm standard error from three repeats ($n = 3$). Standard errors are indicated in figures using the bars.

“MAPK signaling pathway” (8 genes) (Figure 5D). Previous studies observed similar mRNA accumulation related to pathogenesis/disease-related genes and MAPK signaling in Cd-treated roots of tall fescue (Zhu et al., 2018). These results indicate important crosstalk between pathogen stress signaling and HM signaling.

Transcription factors involved in Cd stress response in clove basil

Given that transcription factors (TFs) play important roles in regulating the expression of Cd-responsive genes, and therefore, it is necessary to identify Cd-inducible TFs. A total of 78 TFs that could respond to Cd stress were screened out, mainly belonging to 12 families including *bHLH*, *WRKY*, *AP2/ERF*, and *MYB* (Figure 6A). *bHLH* family members constituted a large proportion (17 out of 78) of Cd-responsive TFs, while a subset of

Zinc finger family members, including the *WRKY* subfamily, were involved in responses to Cd stress. Relatively fewer TF members were characterized as *MYB*, *HSF*, *GRAS*, *PBI*, *MADS*, *MYC*, *VQ*, and *SPL* family members (Figure 6A).

Among the Cd-inducible TF genes, eight showed significantly increased expression after Cd treatment, with seven continually induced by Cd stress. Only one gene, *bHLH160* (c51250.graph_c0), was sharply induced at 24 h in the leaves of clove basil. Five of the seven Cd-inducible TF genes belonged to the *bHLH* family (c58694.graph_c0, c57229.graph_c0, c56571.graph_c1, c51522.graph_c0, and c62303.graph_c0), while one belonged to the *WRKY* family (c63519.graph_c0) and another belonged to the *AP2/ERF* family (c48938.graph_c0) (Figures 6B–I). These results suggest that *bHLH* TFs may be involved in the adaptation regulation of Cd stress, highlighting the importance of these five Cd-inducible *bHLH* genes as potential candidates for improving Cd tolerance in clove basil through genetic engineering.

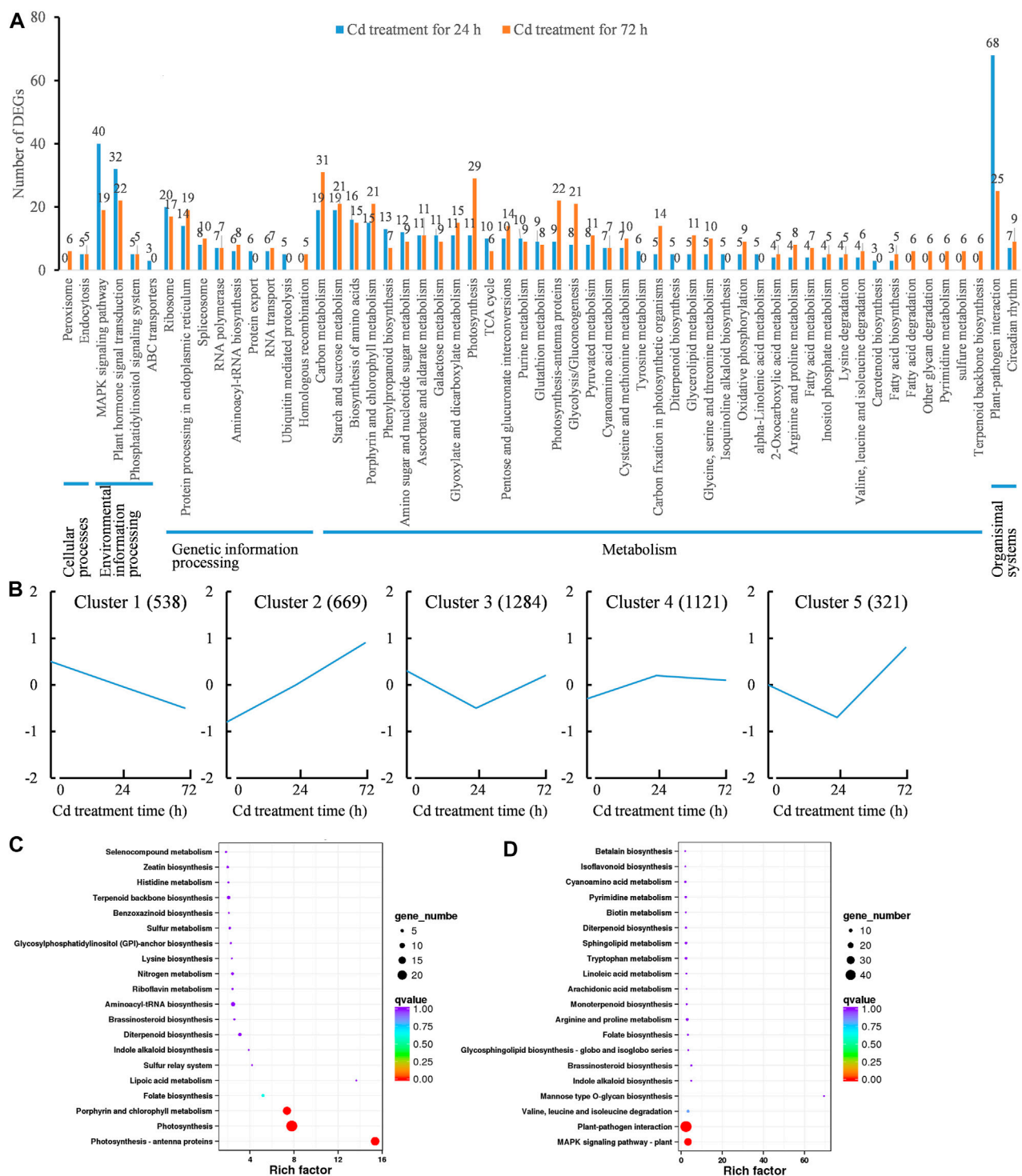


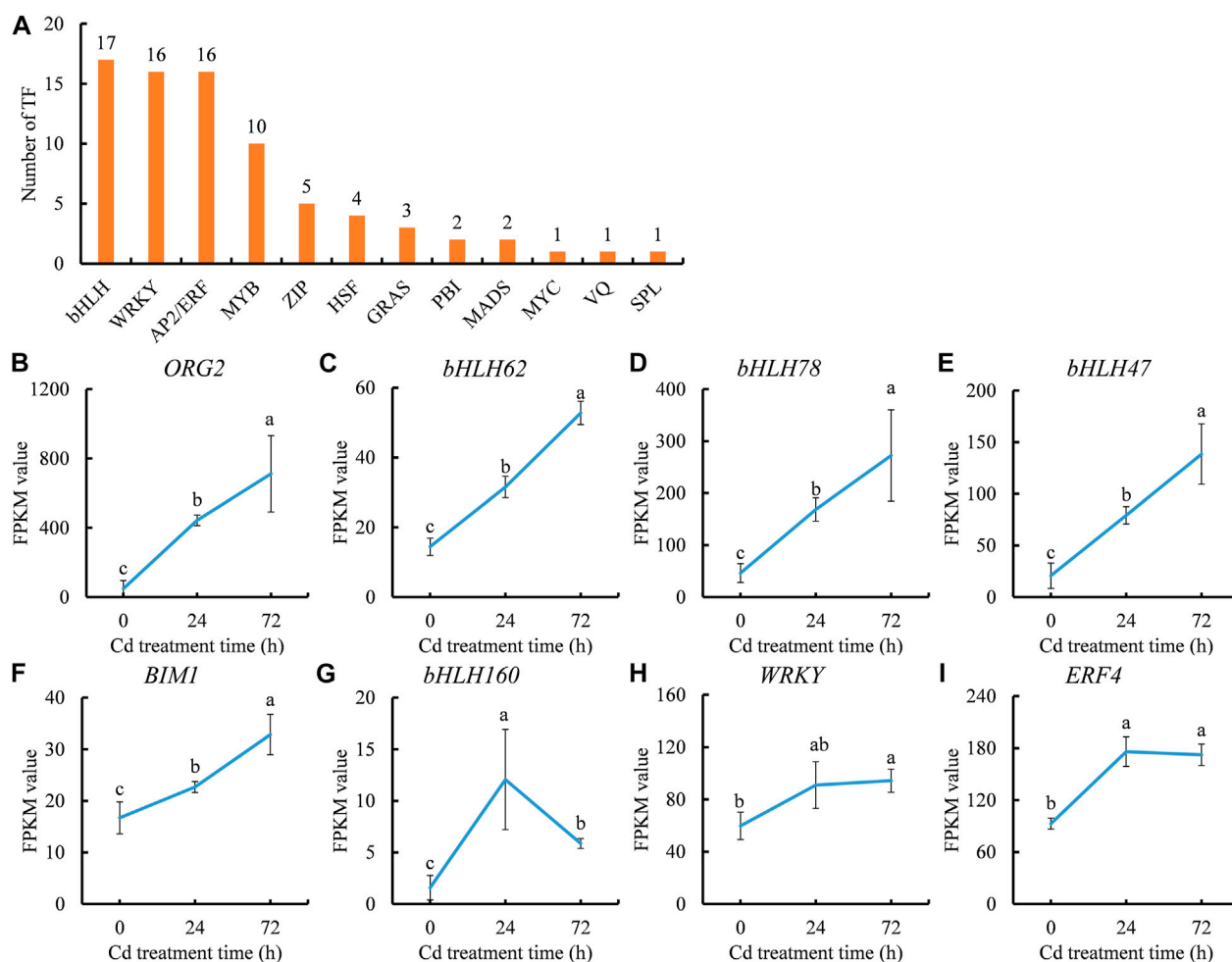
FIGURE 5

Functional annotation and expression patterns of differentially expressed genes (DEGs) caused by Cd treatment (1.6 mg/L). (A), KEGG classification analysis of DEGs. (B), expression patterns of DEGs during 72 h of Cd treatment. (C), KEGG enrichment analysis of DEGs from the Cluster 1; (D), KEGG enrichment analysis of DEGs from the Cluster 2.

Effects of Cd treatment on the antioxidant system in clove basil

Cd toxicity in plants is mainly due to oxidative damages (Ahmad et al., 2021), hence, monitoring physiological indices related to

oxidative stress in clove basil after Cd treatment is important. The contents of TSP, activities of SOD and POD in leaves were decreased after Cd treatment, although SOD activity did not show significant differences before and after Cd treatment (Supplementary Figures S8A–C). After 72 h of Cd treatment,

**FIGURE 6**

Cd stress responsive transcription factor (TF) genes (A) and expression trends of Cd stress inducible TF genes (B–I) during the period of Cd treatment. (B), *ORG2* gene. (C), *bHLH62* gene. (D), *bHLH78* gene. (E), *bHLH47* gene. (F), *BIM1* gene. (G), *bHLH160* gene. (H), *WRKY* gene. (I), *ERF4* gene. Each value is presented as means \pm standard error from three repeats ($n = 3$). Statistical differences ($p \leq 0.05$) between treatments were compared using the SPSS software and indicated using different letters above the bars.

POD activities and TSP contents were reduced by 8.52% and 64.72%, respectively, indicating that Cd stress might have repressed the biosynthesis and accumulation of new proteins or promoted protein degradation. CAT activity showed an overall increasing trend and was enhanced by 119.63% after Cd treatment (Supplementary Figure S8D).

Furthermore, the expression profiles of DEGs encoding antioxidant enzymes such as POD, SOD and CAT were analyzed during Cd treatment (Supplementary Figure S8E). Specific genes in the same family showed a significant expression pattern, suggesting that these antioxidant-related genes intricately regulated enzyme activity at the translation level. The changes in activity of these three antioxidant enzymes were validated at the transcriptional level using several representative genes during Cd stress (Supplementary Figures S8F–K). Cd treatment significantly downregulated the expression of two POD genes (Supplementary Figures 8F, G). The expression levels of two SOD genes did not show significant differences at 0 and 72 h (Supplementary Figures 8H, I). However, Cd treatment significantly induced the expression of two CAT genes

(Supplementary Figures 8J, K), consistent with the increased enzyme activity. These results suggest that H_2O_2 scavenging activity was enhanced after Cd treatment, indicated by the significant increase in CAT activity, which may be a potential mechanism of clove basil adaptation to Cd stress.

The glutathione (GSH) biosynthesis pathway is positively related to phytochelatin (PC) synthesis, which helps in mitigating Cd-induced oxidative stress. Therefore, we analyzed the expression profiles of genes in the GSH pathway. During Cd stress, a total of 10 genes encoding glutathione S-transferase (GST) were differentially expressed. Among them, the expressions of four GSTs (c54510.graph_c0, c51495.graph_c4, c59296.graph_c0 and c51498.graph_c0) were initially reduced at 24 h, but then were strengthened at 72 h, whereas the expression levels (c52100.graph_c0, c45992.graph_c0, and c33411.graph_c0) were constantly increased during Cd treatment. The remaining three GSTs (c50364.graph_c0, c62802.graph_c0, and c55400.graph_c0) showed an overall decreased trend in expression after Cd treatment (Supplementary Figures S9).

Two DEGs encoding phytochelatin synthase (PCS) were identified, and their expression levels were significantly increased after long-term Cd treatment (72 h) compared to 0 h. The expression of a glutathione reductase (GR) gene was downregulated by Cd treatment. Cd treatment significantly regulated the expression of four glutathione peroxidase (GPX) genes in clove basil, with all expressions being repressed. Additionally, two genes involved in the biosynthesis pathway of GSH were significantly induced by Cd stress (Supplementary Figures S9). These results suggest that Cd stress influences GSH biosynthesis, and the increased GSH levels would help alleviate Cd-induced oxidative stress in clove basil.

Discussion

Clove basil has abilities to enrich Cd ions and can tolerate mild Cd stress

Cd is a non-essential element for plant growth and development, and is highly toxic to plants (Shaari et al., 2022). Cd exposure is believed to reduce plant biomass (Halim et al., 2020). As clove basil is a cash plant used for producing essential oil, its biomass positively correlates with the yield of essential oils. Therefore, it is important to determine how much Cd stress intensity influences the growth of clove basil seedlings for practical applications. In this study, we evaluated the effects of different dosages of CdCl₂ treatments on the growth of clove basil and Cd ion contents in clove basil tissues. The results showed that Cd ion contents in stress treatments varied from 4.68 mg/kg in leaves of 0.4 mg/L Cd group to 931.11 mg/kg in roots of 6.5 mg/L Cd group (Figure 2), implying that clove basil seedlings had the ability to accumulate or enrich Cd ions in inner tissues. However, low-level Cd treatment (0.4 mg/L) did not reduce the growth and biomass of clove basil plant (Figure 1). These results demonstrated that clove basil might be an alternative plant for the utilization and remediation of mildly Cd-contaminated soils. The application of clove basil in remediating Cd-contaminated soils can reduce the safety risk of agricultural products while enhancing the utilization efficiency of HM-contaminated soils. To our best knowledge, this study is the first report on investigations of Cd enrichment ability in clove basil.

Cd accumulation in plants shows obvious tissue-specificity, with Cd toxicity showing a dose-effect (Jin et al., 2018). The highest accumulation of Cd ions is in roots followed by shoots and leaves in *Polygonatum sibiricum*, a traditional Chinese medicinal herb (Xie et al., 2021). Our study confirms that in all Cd treatments, leaves had the lowest Cd contents while roots contained the highest Cd contents (Figure 2), which benefits the safe utilization of Cd-contaminated soils, as essential oil is mainly generated and stored in leaves. High levels of Cd stress inhibit plant growth and damage the photosynthesis system of mung bean (Aqeel et al., 2021). Similarly, we observed that the growth performance of clove basil seedlings was significantly reduced by high levels of Cd treatments (0.8, 1.6, 6.5 mg/L), and the chlorophyll pigment contents in clove basil leaves were also reduced following Cd treatments (Figure 1). These results suggest that high degrees of Cd stress negatively affect the growth and biomass of clove basil, which could have implications for the essential oil yield of this important aromatic

plant. Therefore, identification of key Cd-responsive genes that have important roles in the tolerance of Cd stress or in the regulation of Cd response is necessary for molecular breeding of clove basil tolerant to moderate and even severe Cd stress.

Transcriptomic data evaluation and functional analysis of Cd-responsive genes

The above results demonstrate the potential of clove basil to safely utilize and remediate Cd-contaminated soils. However, the lack of available genomic data for clove basil greatly restricts molecular studies of this species (Parasar et al., 2021). The development of nucleic acid sequencing technologies has made it possible to use RNA-seq combined with bioinformatics analysis to reveal plant responses to Cd stress at the molecular level and screen key genes that could positively regulate Cd tolerance in plants. In this study, we sequenced the whole transcriptome of clove basil under Cd stress conditions using Illumina HiSeq technology.

The quality of RNA-seq data is crucial for reliable analysis results. In this study, over 51.00% of transcripts were successfully annotated (Supplementary Table S4), which was higher than other plant species without reference genomic data, such as 40.38% for *Crossostephium chinensis* (Yang et al., 2017), 46.60% for peppermint plant (Wang et al., 2023), and 35.92% for wild paper mulberry (Xu et al., 2019). The expression patterns of representative DEGs from RNA-seq were highly consistent with those detected by RNA-seq (Figure 4). These results suggest that both the quality of transcriptomic data and the mapping effectiveness were high. Sanger sequencing also confirmed the high accuracy of transcriptomic data (Supplementary Figure S5).

Cd stress alters the expression levels of certain genes in plants, known as Cd-responsive genes (Zhao et al., 2021). However, the expression pattern of Cd-responsive genes and the regulation intensity vary between plant species. For example, in *Populus × canadensis* “Neva” leaves, it was reported that 2,816 genes were Cd-responsive, and 1,346 (47.80%) decreased in abundance (Li et al., 2021). In the roots of tall fescue, a total of 2,594 Cd-responsive genes were detected, while only 52 DEGs were found in the leaves (Zhu et al., 2018). Our study found that Cd treatment significantly regulated the expression of over 1,800 DEGs in clove basil leaves, with the numbers of downregulated and upregulated DEGs being almost equal (Figure 3). These results demonstrate that plant responses to Cd stress vary greatly among different species and organs. Investigating the Cd stress response in clove basil would provide new insights for understanding the Cd adaptation mechanisms of plants.

The functions of the DEGs were subsequently annotated through GO, COG, and KEGG classification. In this study, the downregulated DEGs were found to be significantly enriched in the photosynthesis pathway (Figure 5C). Additionally, seedling treated with Cd stress led to a significant reduction in SPAD values and chlorophyll contents in clove leaves (Figure 1). Furthermore, Cd treatment resulted in high Cd accumulations in leaves (Figure 2) and caused severe injuries to young leaves of seedling under Cd stress (Supplementary Figure S3). These results together suggest that Cd stress induces damages to the photosynthesis system, which could explain the observed retardation in the growth of clove basil seedlings. Therefore, implementing some agronomic technologies

to enhance photosynthetic efficiency could be a viable approach to improve Cd resistance in clove basil.

More upregulated DEGs were enriched in “plant-pathogen interaction” and “MAPK signaling pathway” pathways in the KEGG (Figure 5D). The GO analysis showed that DEGs participated in many significant GO terms in response to Cd stress (Supplementary Figure S7). Similar results were observed in *Paspalum vaginatum* Swartz leaves subjected to Cd stress in a former study (Xu et al., 2022). These results suggest that clove basil made systematic and coordinated responses to adapt to Cd stress. Therefore, the functional annotation of Cd-responsive genes could provide key genes for genetically reforming clove basil with strong Cd resistance, which will facilitate the bioremediation and effective utilization of Cd-contaminated lands.

Major TFs and genes induced by Cd treatment and their possible roles in regulating Cd tolerance in clove basil

Transcriptomic analyses in various plant have revealed the involvement of many TFs in the transcriptional regulation of Cd-responsive genes (Zhu et al., 2018; Fasani et al., 2019; Wang et al., 2021b). In this study, identified Cd-responsive TF genes belonged to different families, including bHLH, WRKY, AP2/ERF, MYB, ZIP, HSF, and GRAS (Figure 6A), indicating the complexity of plant responses to Cd stress. Most of these TF genes were repressed by Cd stress in clove basil leaves, demonstrating the inhibitory effects of Cd stress on plants at the transcriptional regulation level. This result aligns with the morpho-physiological data, demonstrating that Cd had a detrimental effect on the growth of clove basil. Another explanation for this result is that the downregulated TF genes might have negative regulatory roles in gene expression during Cd stress adaptation.

In tea plants, transcriptomic analysis showed that many DEGs were produced after 10 or 15 days of Cd exposure, including one gene encoding an ERF protein that was positively correlated with five structural genes, including three *CsGolS* family genes, one *CsNCED*, and one *CsHIPP*, using weighted gene co-expression network analysis (Liu et al., 2023). In *Arabidopsis thaliana*, *AtMYB59* was induced by Cd treatment, and the genes directly and indirectly regulated by *AtMYB59* were mainly involved in calcium homeostasis and signaling through transcriptomic analysis (Fasani et al., 2019). In soybean plants, 26 Cd-responsive WRKY genes were upregulated, while 3 were downregulated by Cd treatment. Among these, *WRKY142* could activate the expression of three Cd tolerance genes, *ATCDT1*, *CDT1-1*, and *CDT1-2*, by directly binding to the W-box element in their promoters (Cai et al., 2020). These reports, together with our results, suggest that Cd-responsive TF genes may play important transcriptional regulatory roles in the expression of HM-stress-related genes.

Recently, several bHLH TFs have been identified as involved in the regulation of Cd tolerance in plants. The ORG3-like gene *GmORG3*, a bHLH family gene, was significantly induced by external Cd stress in soybean plant, and overexpression of *GmORG3* enhanced Cd tolerance and stabilized Fe homeostasis (Xu et al., 2017). Plants overexpressing *bHLH104* exhibited enhanced Cd tolerance in *Arabidopsis thaliana* (Yao et al., 2018).

In this study, among 17 Cd-regulated bHLH family genes, 6 bHLH genes (*c58694.graph_c0*, *c57229.graph_c0*, *c56571.graph_c1*, *c51522.graph_c0*, *c62303.graph_c0*, and *c51250.graph_c0*) showed significantly increased expression levels after Cd treatment (Figures 6B–G). A novel gene encoding a MYB TF was identified in rice seedlings exposed to Cd stress through comparative transcriptomic analysis, which might be a candidate target for generating Cd-resistant plants (Chen et al., 2021). These results suggested that upregulated bHLH TF genes might play crucial regulatory roles in the adaptation of clove basil to Cd stress. These genes could potentially serve as candidate TF genes for developing Cd-resistant clove basil.

The *Lamiaceae* family encompasses a wide variety of plants with biomedical and industrial applications, offering significant potential for medical and daily industries (Uritu et al., 2018). Yet, only limited genetic engineering attempts have been reported in plants belonging to the *Lamiaceae* family. Overexpressing *SmNAC1* enhanced tolerance to high concentrations of Zn ions, and Zn ions were significantly enriched in shoot tissues of *Salvia miltiorrhiza*, a member of the *Lamiaceae* family (Zhu et al., 2019). Treatment with copper sulfate sharply induced the expression of three SOD genes, and the authors suggested that these induced SOD genes were the main contributors during Cu stress (Han et al., 2020). Furthermore, it has been reported that some members of bHLH, WRKY, and MYB TF families in plants of the *Lamiaceae* family participated in the Cd tolerance by regulating the activities of enzymes, such as PCS and antioxidant enzymes (Raza et al., 2020; Fan et al., 2023). Introducing *Perilla frutescens* genes into tobacco developed a new variety *N. tabacum* L. var. ZSY, which significantly enhanced Cd tolerance, reflecting higher SOD and CAT activities and more alkaloids in plant tissues (Wei and Guo, 2023). In a previous study, we investigated transcriptomic changes in peppermint young plants when exposed to different intensities of Cd stresses and identified several key genes, such as *bHLH47* and *bHLH100*, that were significantly induced by Cd stress and had potential roles in Cd resistance acquisition (Wang et al., 2023). However, their regulatory roles in Cd resistance have not been thoroughly investigated. Therefore, there is a necessity for extensive studies on molecular investigations and genetic improvements for plants in the *Lamiaceae* family to confer HM tolerance.

Cd treatment affects the antioxidant activity and induces GSH synthesis in clove basil

The activity of antioxidant enzymes is closely related to Cd tolerance in plants. However, the effects of Cd stress on the activity of antioxidant enzymes differ with plant species, tissues, Cd concentration as well as stress duration. For example, Cd toxicity tends to minimize the activities of SOD and CAT in pea (Sandalio et al., 2001) and SOD activity in *Pistia stratiotes* (Zhao et al., 2023). Cd stress at 14 mg/L significantly enhances the activities of POD, SOD and CAT in both leaves and roots of *Dendrobium officinale* seedlings (Jiang et al., 2020), while Cd stress resulted in a significant decline in the activities of GPX, ascorbate peroxidase and CAT in radish (El-Beltagi et al., 2010). In a previous study, we found that long-term exposure to low Cd concentrations induced antioxidant activity in peppermint

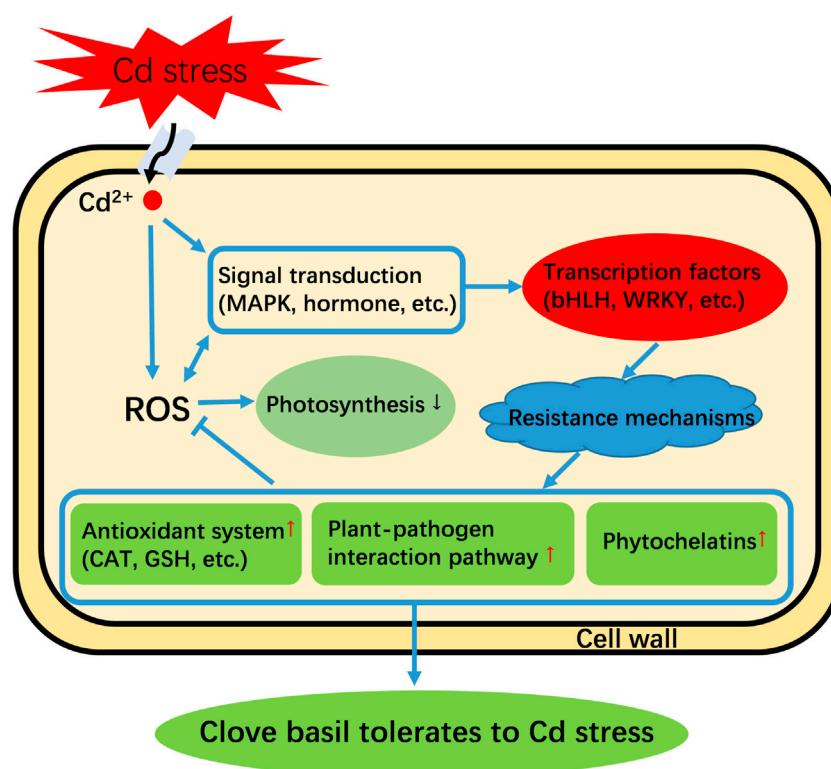


FIGURE 7

An assumed model of Cd tolerance in clove basil. "↑" means this pathway is strengthened. "↓" means this pathway is repressed. "→" means an induction effect. "T" means an inhibitory effect.

leaves, while high-intensity Cd stress inhibited that after long-term exposure (Wang et al., 2023).

In this study, Cd stress did not affect SOD activities, reduced TSP contents and POD activities, but enhanced CAT activities in clove basil (Supplementary Figure S8). These results imply that CAT might play a dominant role in scavenging ROS in clove basil during Cd stress adaptation as only CAT activity was enhanced following Cd treatment. Merely having high CAT activity is insufficient to counteract the Cd-induced ROS burst in clove basil. As a result, seedlings continued to exhibit noticeable harmful effects under Cd stress. The expression patterns of genes encoding these enzymes differed among specific genes, suggesting that Cd stress affected compositions in the antioxidant system at both protein and gene levels in clove basil, and they were complementary to each other in coding for proteins.

GSH is a derivative of glutamic acid, cysteine, and glycine amino acids, which has been considered a ligand that chelates HMs in plants (Al-Khayri et al., 2023). Increased GSH synthesis has been reported in *Solanum nigrum* (Deng et al., 2010), *Bacopa monnieri* (Singh et al., 2006), and *Lepidium sativum* (Gill et al., 2012), with increasing Cd concentrations along with enhanced antioxidant activity. In this study, although we did not measure GSH contents in clove basil after Cd treatment, the expressions of two genes involved in the GSH biosynthesis pathway were significantly upregulated by Cd treatment, while the expressions of genes involved in the process of GSH oxidation were repressed

(Supplementary Figure S9). Therefore, we can reasonably assume that Cd stress induced GSH biosynthesis but inhibited GSH oxidation in clove basil leaves, resulting in the accumulation of reduced GSH substances that improved antioxidant ability to alleviate Cd-induced oxidative stress and provide sufficient GSH for the generation of phytochelatin (PCs).

Cd stress can induce the generation of PCs, which bind Cd and form varied complexes to reduce Cd toxicity in plant cells. The Cys thiolic groups of PCs guard the cytosol from free Cd ions and eventually sequester Cd in the vacuole (Al-Khayri et al., 2023). In this study, we found that Cd stress significantly induced the expression of two PCS genes. Additionally, many GST genes showed significant expression after Cd treatment (Supplementary Figure S9). These results together suggest that Cd treatment promoted the biosynthesis of PCs and indicated the important roles of PCs in reducing the effects of Cd toxicity in clove basil.

Based on the above discussions and expression profiles of genes in response to Cd treatment, we developed a hypothetical model of Cd tolerance in clove basil under Cd stress (Figure 7). When seedlings were subjected to Cd stress, it activated signaling transduction pathways and induced ROS bursts in the cell. The increased levels of ROS cause severe oxidative damage to the photosynthetic system in leaves of clove basil seedlings. The activated signals subsequently regulate the expression of key TF genes, such as *bHLH*, to regulate the expression of genes in stress

defense pathways. Finally, the activated defense system enables clove basil seedlings to endure Cd stress.

Conclusion

Long-term Cd stress caused damage to the photosynthetic and root systems of clove basil, resulting in an insufficient energy supply and reduced performance. A large number of Cd-responsive genes in clove basil were identified, providing valuable resources for genetic and genomic studies of clove basil. More importantly, the results of this study suggest that *bHLH* family TF genes might play crucial roles in the adaptation regulations under Cd stress. Therefore, additional investigations are needed to study the functional roles of these Cd-stress inducible genes, especially TF genes, in the regulation of Cd tolerance in clove basil through transgenic methods in the future. Cd treatment induced CAT activity at both protein and gene levels, suggesting that Cd treatment enhances H₂O₂ scavenging activity in clove basil leaves because CAT is a key antioxidant enzyme that scavenges H₂O₂, and Cd's toxic effects on plants mainly result from triggered ROS bursts by Cd stress. Overall, our study provides new genetic resources for breeding new cultivars with higher Cd resistance, which would greatly facilitate phytoremediation and effective utilization of Cd-polluted soils.

Data availability statement

The datasets presented in the study are deposited in the NCBI Sequence Read Archive (SRA) repository (<https://www.ncbi.nlm.nih.gov/sra>). Accession number PRJNA904532.

Author contributions

BW: Data analysis, Visualization, Writing and editing; YW: Methodology, Data analysis, Writing and editing; XY, YJ, YZ, and XK: Writing and editing; JH: Project administration and

supervision, Writing, editing and review; YX: Conceptualization, Investigation, Plant material management, Writing, editing and review. All authors contributed to the article and approved the submitted version.

Funding

This study was financially supported by the Construction Project for Shaoguan Social Development and Collaborative Innovation of Science and Technology System in 2022 (No. 220607164531948), Guangdong Basic and Applied Basic Research Foundation (No. 2023A1515010494), and Natural Science Project of Shaoguan University (No. SZ2022KJ01).

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/fgene.2023.1224140/full#supplementary-material>

References

- Adewumi, A. J., and Laniyan Dr, T. A. (2020). Contamination, sources and risk assessments of metals in media from Anka artisanal gold mining area, Northwest Nigeria. *Sci. Total Environ.* 718, 137235. doi:10.1016/j.scitotenv.2020.137235
- Agarwal, P., Parida, S. K., Mahto, A., Das, S., Mathew, I. E., Malik, N., et al. (2014). Expanding frontiers in plant transcriptomics in aid of functional genomics and molecular breeding. *Biotechnol. J.* 9, 1480–1492. doi:10.1002/biot.201400063
- Ahmad, P., Raja, V., Ashraf, M., Wijaya, L., Bajguz, A., and Alyemeni, M. N. (2021). Jasmonic acid (JA) and gibberellic acid (GA3) mitigated Cd-toxicity in chickpea plants through restricted Cd uptake and oxidative stress management. *Sci. Rep.* 11, 19768. doi:10.1038/s41598-021-98753-8
- Al-Khayri, J. M., Banadka, A., Rashmi, R., Nagella, P., Alessa, F. M., and Almaghasla, M. I. (2023). Cadmium toxicity in medicinal plants: An overview of the tolerance strategies, biotechnological and omics approaches to alleviate metal stress. *Front. Plant Sci.* 13, 1047410. doi:10.3389/fpls.2022.1047410
- Aqeel, M., Khalid, N., Tufail, A., Ahmad, R. Z., Akhter, M. S., Luqman, M., et al. (2021). Elucidating the distinct interactive impact of cadmium and nickel on growth, photosynthesis, metal-homeostasis, and yield responses of mung bean (*Vigna radiata* L) varieties. *Environ. Sci. Pollut. Res. Int.* 28, 27376–27390. doi:10.1007/s11356-021-12579-5
- Cai, Z. D., Xian, P. Q., Wang, H., Lin, R. B., Lian, T. X., Cheng, Y. B., et al. (2020). Transcription factor GmWRKY142 confers cadmium resistance by up-regulating the cadmium tolerance 1-like genes. *Front. Plant Sci.* 11, 724. doi:10.3389/fpls.2020.00724
- Caicedo-Rivas, G., Salas-Moreno, M., and Marrugo-Negrete, J. (2022). Health risk assessment for human exposure to heavy metals via food consumption in inhabitants of middle basin of the Atrato river in the colombian Pacific. *Int. J. Environ. Res. Public Health.* 20, 435. doi:10.3390/ijerph20010435
- Chen, H. Q., Liang, X. Y., Gong, X. M., Reinfelder, J. R., Chen, H. M., Sun, C. J., et al. (2021). Comparative physiological and transcriptomic analyses illuminate common mechanisms by which silicon alleviates cadmium and arsenic toxicity in rice seedlings. *J. Environ. Sci.* 109, 88–101. doi:10.1016/j.jes.2021.02.030
- Deng, X., Xia, Y., Hu, W., Zhang, H., and Shen, Z. (2010). Cadmium-induced oxidative damage and protective effects of n-acetyl-L-cysteine against cadmium toxicity in *Solanum nigrum* L. *J. Hazard. Mat.* 180 (1–3), 722–729. doi:10.1016/j.jhazmat.2010.04.099
- do Prado, N. B., de Abreu, C. B., Pinho, C. S., Junior, M. M. N., Silva, M. D., Espino, M., et al. (2022). Application of multivariate analysis to assess stress by Cd, Pb and Al in basil (*Ocimum basilicum* L) using caffeic acid, rosmarinic acid, total phenolics, total flavonoids and total dry mass in response. *Food Chem.* 367, 130682. doi:10.1016/j.foodchem.2021.130682

- Dong, Q., Fei, L., Wang, C., Hu, S., and Wang, Z. L. (2019). Cadmium excretion via leaf hydathodes in tall fescue and its phytoremediation potential. *Environ. Pollut.* 252, 1406–1411. doi:10.1016/j.envpol.2019.06.079
- Ekiert, H., Knut, E., Świątkowska, J., Klin, P., Rzepiela, A., Tomczyk, M., et al. (2021). *Artemisia abrotanum* L. (Southern Wormwood)-history, current knowledge on the chemistry, biological activity, traditional use and possible new pharmaceutical and cosmetological applications. *Molecules* 26, 2503. doi:10.3390/molecules26092503
- El-Beltagi, H. S., Mohamed, A. A., and Rashed, M. M. (2010). Response of antioxidative enzymes to cadmium stress in leaves and roots of radish (*Raphanus sativus* L.). *Not. Sci. Biol.* 2, 76–82. doi:10.15835/nsb245395
- Faizan, M., Cheng, S. H., Tonny, S. H., and Robab, M. I. (2022). Specific roles of strigolactones in plant physiology and remediation of heavy metals from contaminated soil. *Plant Physiol. biochem.* 192, 186–195. doi:10.1016/j.plaphy.2022.10.004
- Fan, P., Wu, L., Wang, Q., Wang, Y., Luo, H., Song, J., et al. (2023). Physiological and molecular mechanisms of medicinal plants in response to cadmium stress: Current status and future perspective. *J. Hazard. Mat.* 450, 131008. doi:10.1016/j.jhazmat.2023.131008
- Fasani, E., DalCorso, G., Costa, A., Zenoni, S., and Furini, A. (2019). The *Arabidopsis thaliana* transcription factor MYB59 regulates calcium signalling during plant growth and stress response. *Plant Mol. Biol.* 99, 517–534. doi:10.1007/s11103-019-00833-x
- Gao, J., Sun, L., Yang, X. E., and Liu, J. X. (2013). Transcriptomic analysis of cadmium stress response in the heavy metal hyperaccumulator *Sedum alfredii* Hance. *PLoS One* 8, e64643. doi:10.1371/journal.pone.0064643
- Gautam, M., and Agrawal, M. (2017). Influence of metals on essential oil content and composition of lemongrass (*Cymbopogon citratus* (D.C) Stapf) grown under different levels of red mud in sewage sludge amended soil. *Chemosphere* 175, 315–322. doi:10.1016/j.chemosphere.2017.02.065
- Gill, S. S., Khan, N. A., and Tuteja, N. (2012). Cadmium at high dose perturbs growth, photosynthesis and nitrogen metabolism while at low dose it up regulates sulfur assimilation and antioxidant machinery in garden cress (*Lepidium sativum* L.). *Plant Sci.* 182, 112–120. doi:10.1016/j.plantsci.2011.04.018
- Halim, M. A., Rahman, M. M., Megharaj, M., and Naidu, R. (2020). Cadmium immobilization in the rhizosphere and plant cellular detoxification: Role of plant-growth-promoting rhizobacteria as a sustainable solution. *J. Agric. Food Chem.* 68, 13497–13529. doi:10.1021/acs.jafc.0c04579
- Han, L. M., Hua, W. P., Cao, X. Y., Yan, J. A., Chen, C., and Wang, Z. Z. (2020). Genome-wide identification and expression analysis of the superoxide dismutase (SOD) gene family in *Salvia miltiorrhiza*. *Gene* 742, 144603. doi:10.1016/j.gene.2020.144603
- Jiang, S. J., Sun, J. C., Tong, G. S., Ding, H., Ouyang, J. W., Zhou, Q., et al. (2022). Emerging disposal technologies of harmful phytoextraction biomass (HPB) containing heavy metals: A review. *Chemosphere* 290, 133266. doi:10.1016/j.chemosphere.2021.133266
- Jiang, W., Wu, Z. G., Wang, T., Mantri, N., Huang, H. L., Li, H. W., et al. (2020). Physiological and transcriptomic analyses of cadmium stress response in *Dendrobium officinale* seedling. *Plant Physiol. biochem.* 148, 152–165. doi:10.1016/j.plaphy.2020.01.010
- Jin, C., Nan, Z. R., Wang, H. C., Li, X. L., Zhou, J., Yao, X., et al. (2018). Effect of Cd stress on the bioavailability of Cd and other mineral nutrition elements in broad bean grown in a loess subsoil amended with municipal sludge compost. *Environ. Sci. Pollut. Res. Int.* 25 (8), 7418–7432. doi:10.1007/s11356-017-0994-y
- Li, X., Mao, X. H., Xu, Y. J., Li, Y., Zhao, N., Yao, J. X., et al. (2021). Comparative transcriptomic analysis reveals the coordinated mechanisms of *Populus × canadensis* 'Neva' leaves in response to cadmium stress. *Ecotoxicol. Environ. Saf.* 216, 112179. doi:10.1016/j.ecoenv.2021.112179
- Lichtenthaler, H. K. (1987). [34] Chlorophylls and carotenoids: Pigments of photosynthetic biomembranes. *Meth. Enzym.* 148, 350–382. doi:10.1016/0076-6879(87)48036-1
- Liu, S. Q., Peng, X. Q., Wang, X. J., and Zhuang, W. B. (2023). Transcriptome analysis reveals differentially expressed genes involved in cadmium and arsenic accumulation in tea plant (*Camellia sinensis*). *Plants* 12, 1182. doi:10.3390/plants12051182
- Pandey, J., Verma, R. K., and Singh, S. (2019). Suitability of aromatic plants for phytoremediation of heavy metal contaminated areas: A review. *Int. J. Phytoremediation* 21, 405–418. doi:10.1080/15226514.2018.1540546
- Parasar, N. R., Kumar, R., Natarajan, P., and Parani, M. (2021). *De novo* assembly, annotation and molecular marker identification from the leaf transcriptome of *Ocimum gratissimum* L. *Plant Genet. Resour.* 19, 469–476. doi:10.1017/S1479262121000563
- Raza, A., Habib, M., Kakavand, S. N., Zahid, Z., Zahra, N., Sharif, R., et al. (2020). Phytoremediation of cadmium: Physiological, biochemical, and molecular mechanisms. *Biology* 9 (7), 177. doi:10.3390/biology9070177
- Roosens, N. H., Willems, G., and Saumitou-Laprade, P. (2008). Using *Arabidopsis* to explore zinc tolerance and hyperaccumulation. *Trends Plant Sci.* 13, 208–215. doi:10.1016/j.tplants.2008.02.006
- Saeed, S. T., Khan, A., Kumar, B., Ajayakumar, P. V., and Samad, A. (2014). First report of chilli leaf curl India virus infecting mentha spicata (Neera) in India. *Plant Dis.* 98, 164. doi:10.1094/PDIS-07-13-0750-PDN
- Sandalio, L. M., Dalurzo, H. C., Gómez, M., Romero-Puertas, M. C., and del Río, L. A. (2001). Cadmium-induced changes in the growth and oxidative metabolism of pea plants. *J. Exp. Bot.* 52, 2115–2126. doi:10.1093/jxb/52.364.2115
- Schmittgen, T. D., and Livak, K. J. (2008). A analyzing real-time PCR data by the comparative C(T) method. *Nat. Protoc.* 3, 1101–1108. doi:10.1038/nprot.2008.73
- Shaari, N. E. M., Tajudin, M. T. F. M., Khandaker, M. M., Majrashi, A., Alenazi, M. M., Abdullahi, U. A., et al. (2022). Cadmium toxicity symptoms and uptake mechanism in plants: A review. *Braz. J. Biol.* 84, e252143. doi:10.1590/1519-6984.252143
- Simmer, R. A., and Schnoor, J. L. (2022). Phytoremediation, bioaugmentation, and the plant microbiome. *Environ. Sci. Technol.* 56, 16602–16610. doi:10.1021/acs.est.2c05970
- Singh, S., Eapen, S., and D'Souza, S. F. (2006). Cadmium accumulation and its influence on lipid peroxidation and antioxidative system in an aquatic plant, *Bacopa monnieri* L. *Chemosphere* 62, 233–246. doi:10.1016/j.chemosphere.2005.05.017
- Snyder, J. C., and Desborough, S. L. (1978). Rapid estimation of potato tuber total protein content with coomassie brilliant blue G-250. *Theor. Appl. Genet.* 52, 135–139. doi:10.1007/BF00264747
- Ugbogu, O. C., Emmanuel, O., Agi, G. O., Ibe, C., Ekweogu, C. N., Ude, V. C., et al. (2021). A review on the traditional uses, phytochemistry, and pharmacological activities of clove basil (*Ocimum gratissimum* L.). *Heliyon* 7, e08404. doi:10.1016/j.heliyon.2021.e08404
- Uritu, C. M., Mihai, C. T., Stanciu, G. D., Dodi, G., Alexa-Stratulat, T., Luca, A., et al. (2018). Medicinal plants of the family *Lamiaceae* in pain therapy: A review. *Pain Res. Manag.* 2018, 7801543. doi:10.1155/2018/7801543
- Vashisth, D., Kumar, R., Rastogi, S., Patel, V. K., Kalra, A., Gupta, M. M., et al. (2018). Transcriptome changes induced by abiotic stresses in *Artemisia annua*. *Sci. Rep.* 8, 3423. doi:10.1038/s41598-018-21598-1
- Vilanova, C. M., Coelho, K. P., Luz, T. R. S. A., Silveira, D. P. B., Coutinho, D. F., and de Moura, E. G. (2018). Effect of different water application rates and nitrogen fertilisation on growth and essential oil of clove basil (*Ocimum gratissimum* L.). *Ind. Crop. Prod.* 125, 186–197. doi:10.1016/j.indcrop.2018.08.047
- Wei, K. Q., and Guo, T. T. (2023). Enhancing the potential for cadmium phytoremediation by introducing *Perilla frutescens* genes in tobacco. *Environ. Sci. Pollut. Res. Int.* 30 (27), 70039–70053. doi:10.1007/s11356-023-27392-5
- Wang, B., Lin, L. N., Yuan, X., Zhu, Y. N., Wang, Y. K., Li, D. L., et al. (2023). Low-level cadmium exposure induced hormesis in peppermint young plant by constantly activating antioxidant activity based on physiological and transcriptomic analyses. *Front. Plant Sci.* 14, 1088285. doi:10.3389/fpls.2023.1088285
- Wang, B., Wang, Y. K., Huang, Y. Y., Jiang, Y. Y., He, J. M., and Xiao, Y. H. (2022). Anti-browning effects of citronellal on fresh-cut taro (*Colocasia esculenta*) slices under cold storage condition. *Front. Sustain. Food Syst.* 6, 1001362. doi:10.3389/fsufs.2022.1001362
- Wang, B., Wu, C. S., Wang, G., He, J. M., and Zhu, S. J. (2021a). Transcriptomic analysis reveals a role of phenylpropanoid pathway in the enhancement of chilling tolerance by pre-storage cold acclimation in cucumber fruit. *Sci. Hortic.* 288, 110282. doi:10.1016/j.scienta.2021.110282
- Wang, T. Y., Gao, X. Q., Chen, S. S., Li, D. P., Chen, S. W., Xie, M. H., et al. (2021b). Genome-wide identification and expression analysis of ethylene responsive factor family transcription factors in *Juglans regia*. *Peer J.* 9, e12429. doi:10.7717/peerj.12429
- Xiao, Y. H., Xie, J., Wu, C. S., He, J. M., and Wang, B. (2021). Effects of melatonin treatment on browning alleviation of fresh-cut foods. *J. Food Biochem.* 45 (9), e13798. doi:10.1111/jfbc.13798
- Xiao, Y. H., Zhang, J. L., Jiang, Y. Y., Yuan, Y., Xie, J., He, J. M., et al. (2022). Cinnamic acid treatment reduces the surface browning of fresh-cut taro. *Sci. Hortic.* 291, 110613. doi:10.1016/j.scienta.2021.110613
- Xie, M. D., Chen, W. Q., Dai, H. B., Wang, X. Q., Li, Y., Kang, Y. C., et al. (2021). Cadmium-induced hormesis effect in medicinal herbs improves the efficiency of safe utilization for low cadmium-contaminated farmland soil. *Ecotoxicol. Environ. Saf.* 225, 112724. doi:10.1016/j.ecoenv.2021.112724
- Xu, L., Zheng, Y. Y., Yu, Q., Liu, J., Yang, Z. M., and Chen, Y. (2022). Transcriptome analysis reveals the stress tolerance to and accumulation mechanisms of cadmium in *Paspalum vaginatum* Swartz. *Plants* 11, 2078. doi:10.3390/plants11162078
- Xu, Z. G., Dong, M., Peng, X. Y., Ku, W. Z., Zhao, Y. L., and Yang, G. Y. (2019). New insight into the molecular basis of cadmium stress responses of wild paper mulberry plant by transcriptome analysis. *Ecotoxicol. Environ. Saf.* 171, 301–312. doi:10.1016/j.ecoenv.2018.12.084
- Xu, Z. L., Liu, X. Q., He, X. L., Xu, L., Huang, Y. H., Shao, H. B., et al. (2017). The soybean basic helix-loop-helix transcription factor ORG3-like enhances cadmium tolerance via increased iron and reduced cadmium uptake and transport from roots to shoots. *Front. Plant Sci.* 8, 1098. doi:10.3389/fpls.2017.01098
- Yang, H. Y., Sun, M., Lin, S. J., Guo, Y. H., Yang, Y. J., Zhang, J. X., et al. (2017). Transcriptome analysis of *Crossostephium chinensis* provides insight into the molecular basis of salinity stress responses. *PLoS One* 12, e0187124. doi:10.1371/journal.pone.0187124
- Yao, X. N., Cai, Y. R., Yu, D. Q., and Liang, G. (2018). *bHLH104* confers tolerance to cadmium stress in *Arabidopsis thaliana*. *J. Integr. Plant Biol.* 60, 691–702. doi:10.1111/jipb.12658

- Yuan, X., Tang, B. L., Wang, Y. K., Jiang, Y. Y., He, J. M., Wang, G., et al. (2023). Inhibitory effects of peppermint extracts on the browning of cold-stored fresh-cut taro and the phenolic compounds in extracts. *Front. Sustain. Food Syst.* 7, 1191396. doi:10.3389/fsufs.2023.1191396
- Zhang, C., Tong, C. C., Cao, L., Zheng, P. P., Tang, X. F., Wang, L. H., et al. (2023). Regulatory module WRKY33-ATL31-IRT1 mediates cadmium tolerance in *Arabidopsis*. *Plant Cell Environ.* 46 (5), 1653–1670. doi:10.1111/pce.14558
- Zhao, L., Zhu, Y. H., Wang, M., Ma, L. G., Han, Y. G., Zhang, M. J., et al. (2021). Comparative transcriptome analysis of the hyperaccumulator plant *Phytolacca americana* in response to cadmium stress. *3 Biotech.* 11, 327. doi:10.1007/s13205-021-02865-x
- Zhao, W., Chen, Z. B., Yang, X. Q., Shen, L. Y., Mao, H., and Zhu, S. X. (2023). Integrated transcriptomics and metabolomics reveal key metabolic pathway responses in *Pistia stratiotes* under Cd stress. *J. Hazard. Mat.* 452, 131214. doi:10.1016/j.jhazmat.2023.131214
- Zhu, B., Huo, D. A., Hong, X. X., Guo, J., Peng, T., Liu, J., et al. (2019). The *Salvia miltiorrhiza* NAC transcription factor *SmNAC1* enhances zinc content in transgenic *Arabidopsis*. *Gene* 688, 54–61. doi:10.1016/j.gene.2018.11.076
- Zhu, H. H., Ai, H. L., Cao, L. W., Sui, R., Ye, H. P., Du, D. Y., et al. (2018). Transcriptome analysis providing novel insights for Cd-resistant tall fescue responses to Cd stress. *Ecotoxicol. Environ. Saf.* 160, 349–356. doi:10.1016/j.ecoenv.2018.05.066



OPEN ACCESS

EDITED BY

Rupesh Deshmukh,
Central University of Haryana, India

REVIEWED BY

Santosh Kumar Gupta,
National Institute of Plant Genome
Research (NIPGR), India
Esa Abiso Godana,
Jiangsu University, China

*CORRESPONDENCE

Vinod Chhokar,
✉ vinodchhokar@gmail.com

RECEIVED 07 June 2023

ACCEPTED 10 August 2023

PUBLISHED 07 September 2023

CITATION

Malik C, Dwivedi S, Rabuma T, Kumar R,
Singh N, Kumar A, Yogi R and Chhokar V
(2023), *De novo* sequencing, assembly,
and characterization of *Asparagus*
racemosus transcriptome and analysis of
expression profile of genes involved in
the flavonoid biosynthesis pathway.
Front. Genet. 14:1236517.
doi: 10.3389/fgene.2023.1236517

COPYRIGHT

© 2023 Malik, Dwivedi, Rabuma, Kumar,
Singh, Kumar, Yogi and Chhokar. 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.

De novo sequencing, assembly, and characterization of *Asparagus* *racemosus* transcriptome and analysis of expression profile of genes involved in the flavonoid biosynthesis pathway

Chanchal Malik¹, Sudhanshu Dwivedi¹, Tilahun Rabuma^{1,2},
Ravinder Kumar¹, Nitesh Singh³, Anil Kumar¹, Rajesh Yogi⁴ and
Vinod Chhokar^{1*}

¹Department of Bio and Nano Technology, Guru Jambheshwar University of Science and Technology, Hisar, Haryana, India, ²Department of Biotechnology, College of Natural and Computational Science, Wolkite University, Wolkite, Ethiopia, ³Faculty of Agricultural Sciences, Shree Guru Gobind Singh Tricentenary University, Gurugram, Haryana, India, ⁴UIBT-Biotechnology, Chandigarh University, Mohali, Punjab, India

Asparagus racemosus is known for its diverse content of secondary metabolites, i.e., saponins, alkaloids, and a wide range of flavonoids. Flavonoids, including phenols and polyphenols, have a significant role in plant physiology and are synthesized in several tissues. Despite the diverse role of flavonoids, genetic information is limited for flavonoid biosynthesis pathways in *A. racemosus*. The current study explores full-scale functional genomics information of *A. racemosus* by *de novo* transcriptome sequencing using Illumina paired-end sequencing technology to elucidate the genes involved in flavonoid biosynthesis pathways. The *de novo* assembly of high-quality paired-end reads resulted in ~2.3 million high-quality reads with a pooled transcript of 45,647 comprising ~76 Mb transcriptome with a mean length (bp) of 1,674 and N50 of 1,868bp. Furthermore, the coding sequence (CDS) prediction analysis from 45,647 pooled transcripts resulted in 45,444 CDS with a total length and mean length of 76,398,686 and 1,674, respectively. The Gene Ontology (GO) analysis resulted in a high number of CDSs assigned to 25,342 GO terms, which grouped the predicted CDS into three main domains, i.e., Biological Process (19,550), Molecular Function (19,873), and Cellular Component (14,577). The Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database was used to categorize 6,353 CDS into 25 distinct biological pathway categories, in which the majority of mapped CDS were shown to be related to translation (645), followed by signal transduction (532), carbohydrate metabolism (524), folding, sorting, and degradation (522). Among these, only ~64 and 14 CDSs were found to be involved in the phenylpropanoid and flavonoid biosynthesis pathways, respectively. Quantitative Real-time PCR was used to check the expression profile of fourteen potential flavonoid biosynthesis pathway genes. The qRT-PCR analysis result matches the transcriptome sequence data validating the Illumina sequence results. Moreover, a large number of genes associated with the flavonoids biosynthesis pathway were found to be upregulated under the induction of methyl jasmonate. The present-day study on transcriptome sequence data of *A. racemosus* can be utilized for characterizing genes involved in flavonoid biosynthesis pathways and for functional genomics analysis in *A. racemosus* using the reverse genetics approach (CRISPR/Cas9 technology).

KEYWORDS

Asparagus racemosus, gene ontology, secondary metabolism, real time PCR, gene expression

1 Introduction

Asparagus racemosus, commonly called Shatavari, belongs to the family of Liliaceae, found in low altitudes throughout India (Rani et al., 2022). It is an extensively scandent spinous, branched under-shrub. Leaves are reduced to small, and roots are numerous and fusiform succulent, and it looms in clusters from the basal end of the stem and has a tuberous shape with a diameter ranging from 0.5 to 1.5 cm. (Bopana and Saxena, 2008). Different parts of the plant exhibit different therapeutic and pharmacological properties, but roots are the most widely used as medicine and health tonic. Roots are rich in various steroids, flavonoids, alkaloids, polyphenols, vitamins, and shatavarins and are used as medicine for different ailments, such as weakness, infertility, libido, menstrual irregularity, dyspepsia, hepatitis, allergies (Prakash and Singh, 2006; Chawla et al., 2011; Kumar et al., 2011), and galactagogue, etc. (Gupta and Shaw, 2011). The dried root of the plant is used as a drug treating ulcers and has been identified to control AIDS (Alok et al., 2013).

Moreover, it is commonly used in traditional Ayurveda preparations, especially in female and rejuvenating tonics (Sharma and Bhatnagar, 2010). The complexity of biological functions arises from intricate interactions among multiple components, including the genome, gene products, and metabolites. To comprehend the molecular underpinnings of these intricate interactions, functional genomics approaches, which utilize NGS data, are promising in discovering the molecular mechanisms that drive complex biological processes involving interactions between gene products, metabolites, and the genome. With the advent of next-generation sequencing technology, genomics research has revolutionized, allowing for the efficient, cost-effective sequencing of genes. This has resulted in the identification of new genes associated with metabolic pathways, especially in non-model plants like eucalyptus (Mizrachi et al., 2010), *American ginseng* (Sun et al., 2010), rubber tree (Liu et al., 2015), *Aloe vera* (Choudhri et al., 2018) and many others that lack a reference genome.

Asparagus racemosus contains several secondary metabolites such as flavonoids, alkaloids, and saponins (Hayes et al., 2008; Hossain et al., 2012; Kumar et al., 2016), thereby increasing pharmaceuticals' attention due to its multiple applications. Secondary metabolites such as quercetin, rutin, and kaempferol are the mainly reported flavonoid components in *A. racemosus*. They have numerous therapeutic activities, including anti oxidative action (Taepongsorat and Rattana, 2018), anti-inflammatory, antitumor effects (Mittra et al., 2012), and antimicrobial activities (Aggarwal et al., 2013). Flavonoids, a class of phenylpropanoids widely distributed in the plant kingdom, comprise one of the major secondary metabolite groups with a C6-C3-C6 general structural backbone. The biosynthesis of flavonoid backbone has been extensively reported in natural product chemistry and molecular biology (Ververidis et al., 2007). Irrespective of their structural variations, the origin of all flavonoid compounds can be traced back to the phenylpropanoid pathway, which converts phenylalanine into 4-coumaroyl-CoA. This is the

crucial step in the flavonoid biosynthesis pathway. Chalcone synthase (CHS) is the first enzyme involved in condensation, and chalcone is formed after subsequent intramolecular cyclization of one p-coumaroyl-CoA with malonyl-CoA molecules (Figure 1). Further, chalcone isomerase (CHI) catalyzes the isomerization of chalcone into flavanone. After that, flavone synthase (FS) converts flavanone into flavone. Dihydroflavonol is formed by the 3-hydroxylation of flavanone by the Flavanone 3-hydroxylase enzyme (F3H) (Ferrer et al., 2008). However, various products, such as anthocyanins, rutin, quercetin, flavonols, etc., require many steps, and sequential modifications among them are acylation, methylation, and glycosylation to build the end product. Most of these steps remain to be elucidated in plants in general and *A. racemosus* in particular.

Despite the importance of flavonoids, the genomic information related to their biosynthesis in *A. racemosus* is poorly understood. Therefore, the present study performed a *de novo* transcriptome sequencing and data assembling analysis of *A. racemosus* to elucidate the genes involved in flavonoid biosynthesis pathways and validate the transcriptomic sequencing data using qRT-PCR. Transcriptome sequencing and analysis of *A. racemosus* can pave the way for a wide array of basic and applied research. Sequencing the transcriptome of this plant will provide a comprehensive catalog of the expressed genes and transcripts, providing researchers with an extensive pool of genomic resources. This information can be used to study the molecular and biochemical processes that underlie the various physiological functions of *A. racemosus*. The application of transcriptome sequencing and analysis in *A. racemosus* research can offer a wealth of genomic resources and insights that can be utilized in various fields of study, including agriculture, medicine, and biotechnology.

2 Methods

2.1 Sample preparation and total RNA isolation

Two-year-old plantlets (*A. racemosus*) regenerated via clonal propagation (Kumar Kar and Sen, 1985) were obtained from the herbal garden of Chaudhary Charan Singh Haryana Agricultural University, Hisar, Haryana, India. Plants were cultivated in a pot that contained a mixture of sterilized soil, sand, and compost in a 2:1:1 ratio under greenhouse conditions at 24°C–28°C and a 16-h light/8-h night cycle. Fresh leaves and roots were harvested from a healthy plant, splashed frozen in liquid nitrogen, and stored at –80°C for further use. The total RNA was extracted from the root and leaves tissues using TRIzol® Reagent (Thermo Fisher Scientific) method. To ensure the presence of the 28S and 18S bands, the total RNA's quality was assessed on a 1% denatured agarose gel at 100 V for 30 min (Supplementary Figure S1). Further, the total RNA quality and quantity were analyzed using a Qubit fluorometer. Finally, the quality of RNA was evaluated using Bioanalyzer Agilent DNA HS Chip (Thermo fisher scientific).

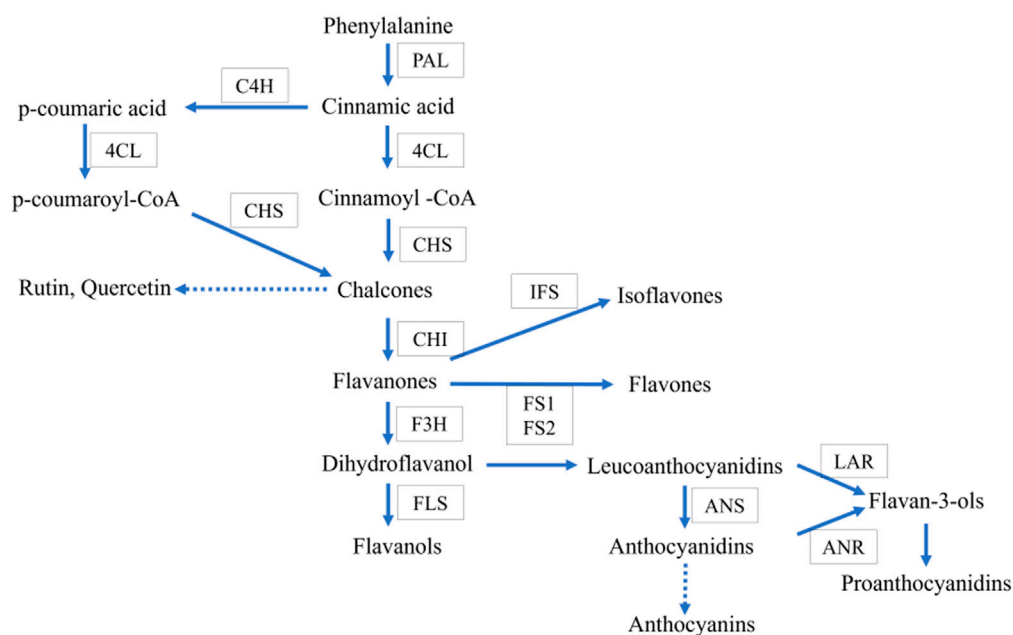


FIGURE 1

Proposed common flavonoid biosynthesis. The enzymes catalyzing each step are in boxes Phenylalanine ammonia-lyase (PAL), Chalcone synthase (CHS), cinnamate-4- hydroxylase (C4H), chalcone isomerase (CHI), flavone synthase (FS), flavonoid-3- hydroxylase (F3H), Flavonol synthase (FLS), anthocyanidin synthase (ANS), isoflavon synthase (IFS).

2.2 Library preparation

The pair-end cDNA sequencing library was prepared by pooling 4 µg total RNA of root and leaf using TrueSeq® Stranded mRNA sample preparation kit (Illumina) as per the manufacturer's instruction. The total RNA isolated from the root and leaf was pooled in an equimolar ratio. Poly- T-attached magnetic beads were utilized to enrich mRNA from the total RNA, which was then enzymatically fragmented and converted into the first-strand cDNA. Subsequently, a second-strand mix was employed to facilitate RNA-dependent synthesis for synthesizing the second strand cDNA from the first cDNA strands. The double-stranded cDNA was purified using AMPure XP beads (Agencourt Biosciences). In the double-strand cDNAs, end repair, A-tailing, and adapter ligation were performed and finally ended with 18 cycles of PCR amplification of the adaptor-ligated library. Following the manufacturer's instructions, the Tape Station 4200 (Agilent Technologies) and the High Sensitivity (HS) D5000 Screen Tape assay kit was utilized to analyze the PCR amplified library.

2.3 Illumina RNA sequencing and data trimming

For cluster generation and sequencing cDNA pair-end library prepared from the pooled sample was loaded onto Next Seq500. The library sequencing and raw data generation for both samples were performed by Xcelris Genomics, using the Illumina Hi-Seq platform with 2 × 150 bp chemistry. The raw sequencing data produced underwent filtration using Trimmomatic v0.30, which involved applying several parameters. These parameters included cutting the read once the average quality within the window falls below

a threshold of 25, cutting the bases off at the beginning of the read if they are below a quality threshold of 25, cutting the bases off at the end of the read if they are below a quality threshold of 25, and discarding any reads that are below a length of 50 bp.

2.4 De novo transcriptome assembly, validation, and CDS prediction

Following the removal of low-quality reads, the remaining high-quality reads were combined and assembled into transcripts using RNA-Seq Assembler Trinity, utilizing a Kmer 25. The validation of transcript assembly was conducted using the CLC Genomics workbench, as depicted in Figure 2. To eliminate isoforms produced during assembly and obtain unigenes that could no longer be extended, non-redundant transcripts were further clustered using CDHIT-EST-454 with 95% identity and query coverage. TransDecoder-v5.3.0 was used to predict coding sequences from unigenes, using specific criteria such as the presence of an open reading frame (ORF) with a minimum length and a log-likelihood score >0.

The coding score was most significant when the ORF was scored in the first reading frame compared to the other five frames. Subsequently, downstream analysis was conducted using CDS with 80% coverage and 3X read depth.

2.5 Gene ontology analysis

To annotate the predicted CDS, a search was conducted against NCBI (<http://www.ncbi.nlm.nih.gov>) using the Basic Local

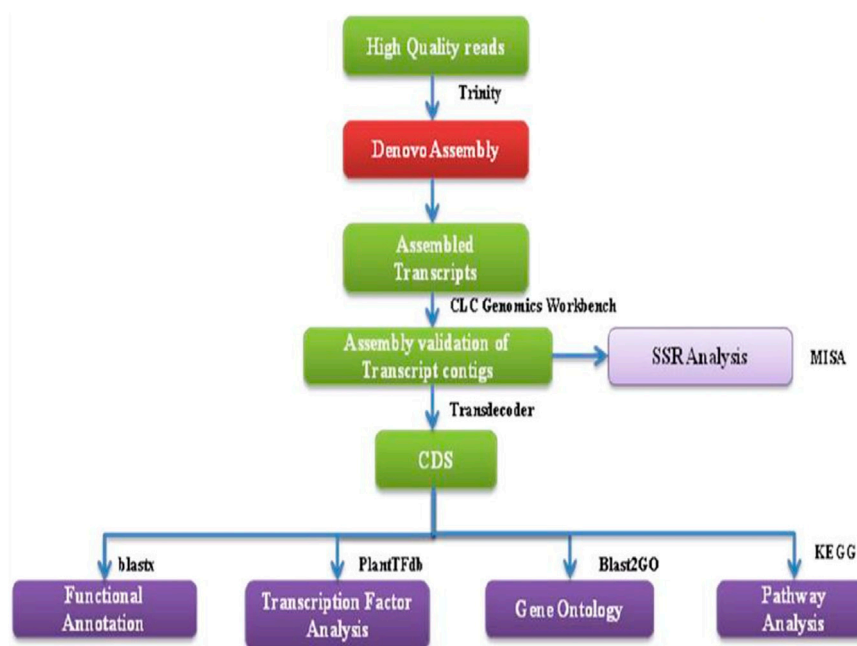


FIGURE 2

A flowchart showing the bioinformatics workflow of *de novo* transcriptome sequencing strategy and analysis for NGS-Asparagus leaf and root sample.

Alignment Search Tool (BLASTx) with a significance threshold cut-off of E-value $\leq 1e-05$. Gene ontology prediction was carried out using the Blast2GO (v 2.7.0) program. The predicted CDS were further categorized into three domains representing gene product properties, namely: Biological Process (BP), Molecular Function (MF), and Cellular Component (CC), using gene ontology mapping analysis.

2.6 Functional annotation of KEGG pathways

To identify the predicted CDS's possible involvement, the CDS was mapped into the reference canonical pathways Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Kanehisa et al., 2004). The potential involvement of predicted CDS was categorized into 25 KEGG pathways under five significant categories, i.e., metabolism, genetic information processing, environmental information processing, cellular processing, and organismal system. The information obtained upon the KEGG pathways analysis included KEGG Orthology (KO) assignments, their corresponding enzyme commission (EC) number, and the prediction of the metabolic pathways using the KEGG automated annotation server KASS (Moriya et al., 2007).

2.7 Transcription factor analysis

Transcription factor identification was performed using P2TF (Predicted Prokaryotic Transcription Factors), which is an integrated and comprehensive database relating to transcription factor proteins. All the predicted CDS were searched against the

Plant transcription factor database (PlantTFdb) using BLASTX to obtain the transcription factor from pooled leaf and root CDS. The transcription factors were identified within the coding regions.

2.8 SSR identification

The observed unigenes were assembled, and SSRs determined using the MicroSatellite Identification tool (MISA, <http://pgrc.inpkn-gatersleben.de/misa/>). A minimum unit size cut-off of 6 was used to report a dinucleotide repeat; for a trinucleotide repeat size cut-off was 4, and for SSRs of sizes 4, 5, and 6 repeats, the cut-off size was 3. A maximum distance of 100 nucleotides was allowed between the two SSRs.

2.9 Methyl jasmonate treatment

The plantlets of control and methyl jasmonate-treated plants were grown in three biological replications. Methyl jasmonate treatment was performed with the foliar spray containing an aqueous solution of 250 μ M MeJA and 0.1% Triton X-100 (Ku and Juvik, 2013).

2.10 Gene validation analysis using qRT-PCR

For the validation of the transcriptomic sequencing data by real-time quantitative PCR (qRT-PCR), fourteen identified genes involved in the flavonoid biosynthesis pathways were selected from *A. racemosus* transcriptome sequencing data, and primers

TABLE 1 Selected genes involved in biosynthesis of flavonoid.

Sr No.	Gene name	EC no.	Unique sequence
1	Trans cinnamate-4-monooxygenase	1.14.13.11	cds_2444
2	Anthocyanidin-3-O-glycosyl transferase	2.4.1.115	cds_12314
3	Anthocyanidin glycosyltransferase	2.4.1.115	cds_12317
4	Shikimate hydroxyl cinnamoyl transferase	2.3.1.133	cds_14257
5	Shikimate hydroxyl cinnamoyl transferase	2.3.1.133	cds_14258
6	Isoflavanone -2- hydroxylase	1.14.13.52	cds_14383
7	Flavanol synthase	1.14.11.23	cds_14436
8	Flavanol synthase	1.14.11.23	cds_14711
9	Chalcone synthase	2.3.1.74	cds_15823
10	Caffeoyl CoA methyltransferase	2.1.1.104	cds_16000
11	Naringenine-3- dioxxygenase	1.14.11.9	cds_16714
12	Quemoyl quinate monooxygenase	1.14.13.36	cds_17418
13	Flavonol monooxygenase	1.14.13.21	cds_22665
14	Flavonol glycosyltransferase	2.4.1.91	cds_34397

were designed for the fourteen candidate genes, and housekeeping genes (GAPDH) using Primer Express software (version 3.0.1) (Table 1). The total RNA was isolated using the CIA-PCIA method (Choudhri et al., 2018) from the root and leaf samples and plants externally treated with methyl jasmonate (250 μ M). According to the manufacturer's protocol, the isolated RNA (1 μ g) was converted to cDNA using the PrimeScript first strand cDNA Synthesis Kit. The qRT-PCR quantification was performed using the Step One Real-Time PCR Instrument (Applied Biosystems) and Takara TB Green™ Premix Ex Taq™ II (Tli RNaseH Plus). The PCR was conducted in a 20 μ L volume containing 4 μ L diluted cDNA, 250 nM forward primer, 250 nM reverse primer, and 10 μ L TB Green Premix Ex Taq II (Tli RNaseH Plus) (2X) (Takara) using the following conditions: 95°C for 3 min, 40 cycles of 95°C for 15 s, 58°C for 15 s, 72°C for 20 s and final extension at 72°C for 5 min. Three technical replications were used for the qRT-PCR reaction of each biological sample. The housekeeping gene GAPDH was utilized as a reference gene in qRT-PCR gene validation analysis for data normalization (Jian et al., 2008). The relative expression level was determined using the $2^{-\Delta\Delta CT}$ method (CT Target- CT GAPDH) (Schmittgen and Livak, 2008) based on three technical replicates of each sample.

3 Results

3.1 De novo transcriptome assembly and validation

The quality of total RNA was checked on agarose gel-electrophoresis and utilized for *de novo* transcriptome Illumina sequencing analysis (Supplementary Figure S1). Illumina Hi-Seq Platform sequencing results of cDNAs prepared from total RNA resulted in 23,101,000 high-quality reads (2 × 150 bp;

6,667,619,278 nucleotides) with paired-end raw reads (~6.6 GB) from the pooled assembled transcript of roots and leaf samples of *A. racemosus* (Table 2). The pooled assembly from the two libraries generated 45,647 assembled transcripts comprising 76,398,686bp total transcriptome length with a mean transcript length of 1,674 and N50 of 1,868bp (Table 2). CDS prediction analysis from 45,647 pooled assembled transcripts resulted in 45,444 CDS with a total length and mean length of 76,398,686 and 1,674 bp, respectively (Table 2). Transcripts distribution analysis revealed the highest number of transcripts (12,324) above 2000bp length, followed by 7,466 transcripts in the length range of 800-1,000bp (Table 2; Figure 3). Low numbers of transcripts (2203) were observed with a 600-800bp length range of transcripts. Moreover, the highest number of CDS (11,656) predicted having a CDS length greater than 1200bp, followed by CDS (8979), which had a length distribution in the range of 400-600 bp.

3.2 Functional annotation and GO classification distribution

For functional gene analysis, the identified CDS were searched against the NCBI non-redundant (NR) protein database, resulting in the annotation of 42,396 CDS. Among the 45,444 CDS identified, a total of 3,048 CDS were categorized as novel, indicating that they were detected in the RNA-Seq data without any significant hits (Table 2). The majority of the hits were found to be against *Vitis vinifera* (20%), followed by *Oryza sativa* (15%), *Theobroma cacao* (10%), and *Setaria italica* (7%) (Figure 4A). GO mapping was carried out to assign the function for BLASTX annotated CDS using the Blast2GO program. Accordingly, 19,550 CDS were found to be involved in biological processes, 19,873 in molecular function, and 14,577 in cellular components (Table 2; Figure 4). The

TABLE 2 Statistical analysis of whole transcriptome data of *A. racemosus* under methyl jasmonate stress treatment.

Parameters	Asparagus pooled sampled
High-quality reads	23,101,000
Total Number of bases	6,667,619,278
Total data (in Gb)	6.6
Pooled assembly statistics	
Transcript description	Transcript stat
Total transcripts	45,647
Total transcript length (bp)	76,398,686
N50	1,868
Maximum transcript length (bp)	7,847
Mean Transcript Length (bp)	1,674
Length range of transcript	No. of transcript
600 < transcript ≤800	2203
800 < transcript ≤1,000	7,466
1,000 < transcript ≤1,200	6352
1,200 < transcript ≤1,400	5330
1,400 < transcript ≤1,600	4611
1,600 < transcript ≤1800	3977
1800 < transcript ≤2000	3384
Transcript >2000	12,324
Length of CDS	No. of CDS
200 < CDS ≤400	6421
400 < CDS ≤600	8979
600 < CDS ≤800	7,421
800 < CDS ≤1,000	6272
1,000 < CDS ≤1,200	4695
>1,200	11,656
The CDS pooled Data Distribution Statistics	
Sample Name	Pooled CDS
Total no. of CDS	45,444
# CDS with Blast Hits	42,396
# CDS without Blast Hits	3,048
Domains	No.CDS
Biological process	19,550
Molecular function	19,873
Cellular Component	14,577

highest number of CDS (>90%) were involved in binding (11,179) (in the molecular function), metabolic process (4763 CDS) (in biological process) and membrane (6961 CDS) (in the cellular component) (Figure 4B). Some other genes were also identified,

including transporter (1541 CDS) and signaling (356 CDS) were involved in biological process and cell (1179 CDS) and envelope (702 CDS) involved in cellular component. Moreover, in molecular function category, the GO distribution assigned CDS related to enzyme encoding involved in the flavonoid's biosynthetic pathways such as *phenylalanine ammonia-lyase* (13) and *chalcone isomerase* (6).

3.3 Transcription factor analysis

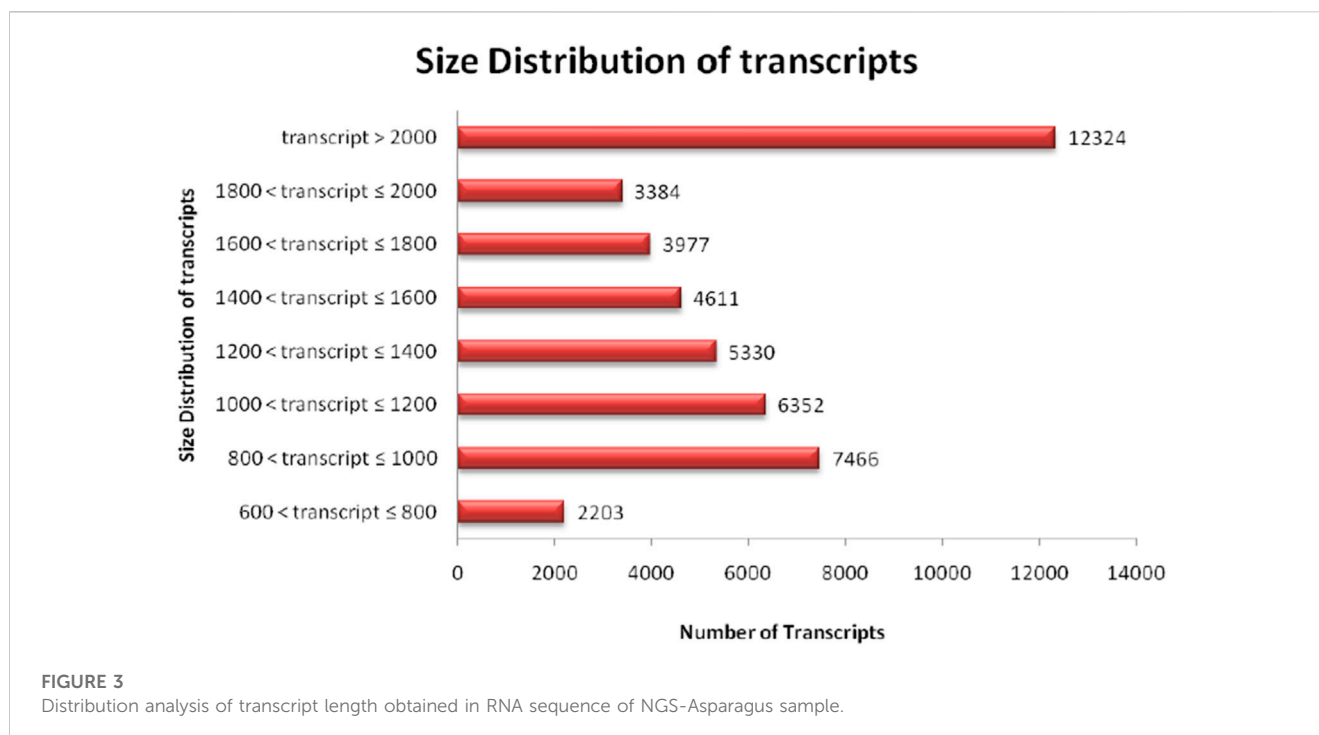
Transcription factors (TFs) are proteins that bind to DNA and control gene expression by activating or inhibiting transcription. Transcription factor analysis generated 80 different transcription factor families in the NGS-Asparagus sample. Transcription factor analysis obtained the highest CDS (905) assigned to Cysteine3Histidine (C3H), followed by a far-red-impaired response (FAR1) (638 CDS) and NAC transcription factors (612 CDS) and bHLH (567) (Figure 7). The NAC family of plant transcription factors is well-known for its involvement in plant stress responses. Additionally, the RNA-Seq data analysis revealed the presence of 454 CDS related to MYB-related transcription factors, which play crucial roles in plant stress responses as well as other biological processes such as development, differentiation, and defense metabolism (Choudhri et al., 2018). Other transcription factors identified were mitochondrial transcription termination factor (mTERF) (206), Tumor necrosis factor receptor-associated factors (TRAFs) 243 (222), Auxin/indole-3-acetic acid (Aux/IAA) (230), MADS (524), C2H2 (408), WRKY (269), SNF2 (294), AP2-EREBP (262), RWP-RK (239), HB (235), FHA (230), etc.

3.4 KEGG pathway mapping of CDS

The identified CDS were mapped to the reference canonical pathways in KEGG to analyze the involvements of predicted CDS in a particular pathway. Accordingly, the KEGG pathway mapping analysis revealed that 6,353 CDS were enriched in 25 different KEGG pathway categories. A large number of CDS (3,065) were mapped into metabolic pathways, the majority of which were shown to be related to carbohydrate metabolism (524), followed by carbon metabolism (382), amino acid metabolism (381), energy metabolism (340) and lipid metabolism (302) (Table 4). A total of 1,561 CDSs were functionally assigned to genetic information processing, among which 645 CDS were involved in translation, followed by folding, sorting, and degradation (522 CDS) and transcription (394 CDS). Moreover, the KEGG pathway mapping functionally assigned a total of 818 CDS involved in environmental information processing with the highest number (532 CDS involved in signal transduction), 770 CDS assigned to cellular processing with the highest number (323 CDS involved in transport and catabolism), and organismal system (139 CDS involved in environmental adaptation).

3.4.1 Potential CDS identification related to secondary metabolism from KEGG mapping

From the *A. racemosus de novo* transcriptome sequencing data, a KEGG pathway mapping functionally assigned many CDS involved



in the secondary metabolism, i.e., metabolism of terpenoids and polyketides (142 CDS), biosynthesis of other secondary metabolites (117 CDS) and xenobiotics biodegradation and metabolism (65). In the current study, several genes related to phenylpropanoid biosynthesis, flavonoid biosynthesis, and metabolism of terpenoids and polyketides have been identified from *A. racemosus* tissues by KEGG Pathway functional annotation. Among 117 CDSs related to the biosynthesis of other secondary metabolites, CDSs were found to be functionally involved in phenylpropanoid (64), flavonoid biosynthesis pathways (12), anthocyanin biosynthesis (2), isoflavonoid biosynthesis (1), flavone and flavonol biosynthesis (3), respectively (Table 4; Figure 4; Figure 5; Figure 6). An organ and developmentally specific pattern of metabolites is created by a combination of reductases, oxygenases, and transferases, amplifying the resulting hydroxycinnamic acids and esters in several cascades based on the few intermediates of the shikimate pathway as the core unit, generated by the general phenylpropanoid metabolism.

3.5 Mining of SSR marker associated with flavonoid pathway genes

Simple sequence repeats (SSRs) or microsatellites are tandem repeats of nucleotide motifs of sizes one to six bp and are highly polymorphic with a ubiquitous presence in all the known genomes. The simple sequence repeats (SSRs) or micro-satellite associated with flavonoids were picked using MISA v1.0 from the pooled transcript, resulting in 7,558 SSRs markers. Among 7,558 SSRs markers identified, the highest number of SSR were 3,810, identified as di-nucleotides, followed by trinucleotide with 3,486 SSRs counts (Table 3). Therefore, the di-nucleotide was the most abundant, with a

frequency of 50.41%, followed by trinucleotide, with a frequency distribution of 46.12%. The SSR sequences are further filtered and validated by removing terminal SSRs and retaining SSRs with a flanking region of 150bp, resulting in 2917 SSRs predicted to contain a flanking region.

3.6 qRT-PCR analysis of flavonoid biosynthesis pathway-related genes

The total RNA extracted with three replications each from *A. racemosus* leaf and MeJA-treated leaf samples was observed as 28S and 18S bands on 1% denaturing agarose gel (Supplementary Figure S2). The Nano-drop spectrophotometer analysis resulted in 1,261 ng/μL (1), 1,358 ng/μL (2), 1,193 ng/μL (3), 1,497 ng/μL (4), 1,948 ng/μL (5) and 1,584 ng/μL (6) concentrations, respectively. Fourteen genes involved in flavonoid biosynthesis pathways were analyzed for their differential expression from pooled (root and leaf) and methyl jasmonate-treated leaf samples. The qRT-PCR analysis revealed a high-fold expression obtained in the gene encoding anthocyanidin-3-o-glycosyl transferase (FC = 15.029), followed by caffeoyl CoA methyltransferase (FC = 12.63) and quorum quinate monooxygenase (FC = 9.55) (Figure 8). Comparatively, a high level of differential gene expression was observed in MeJA-induced expression than in the non-treated leaf sample. A similar expression pattern was observed across all genes involved in the flavonoid biosynthesis pathways in the non-treated leaf samples.

4 Discussion

Asparagus racemosus is a Rasayana medicine that has been used since the ancient period to treat different diseases. In India, the

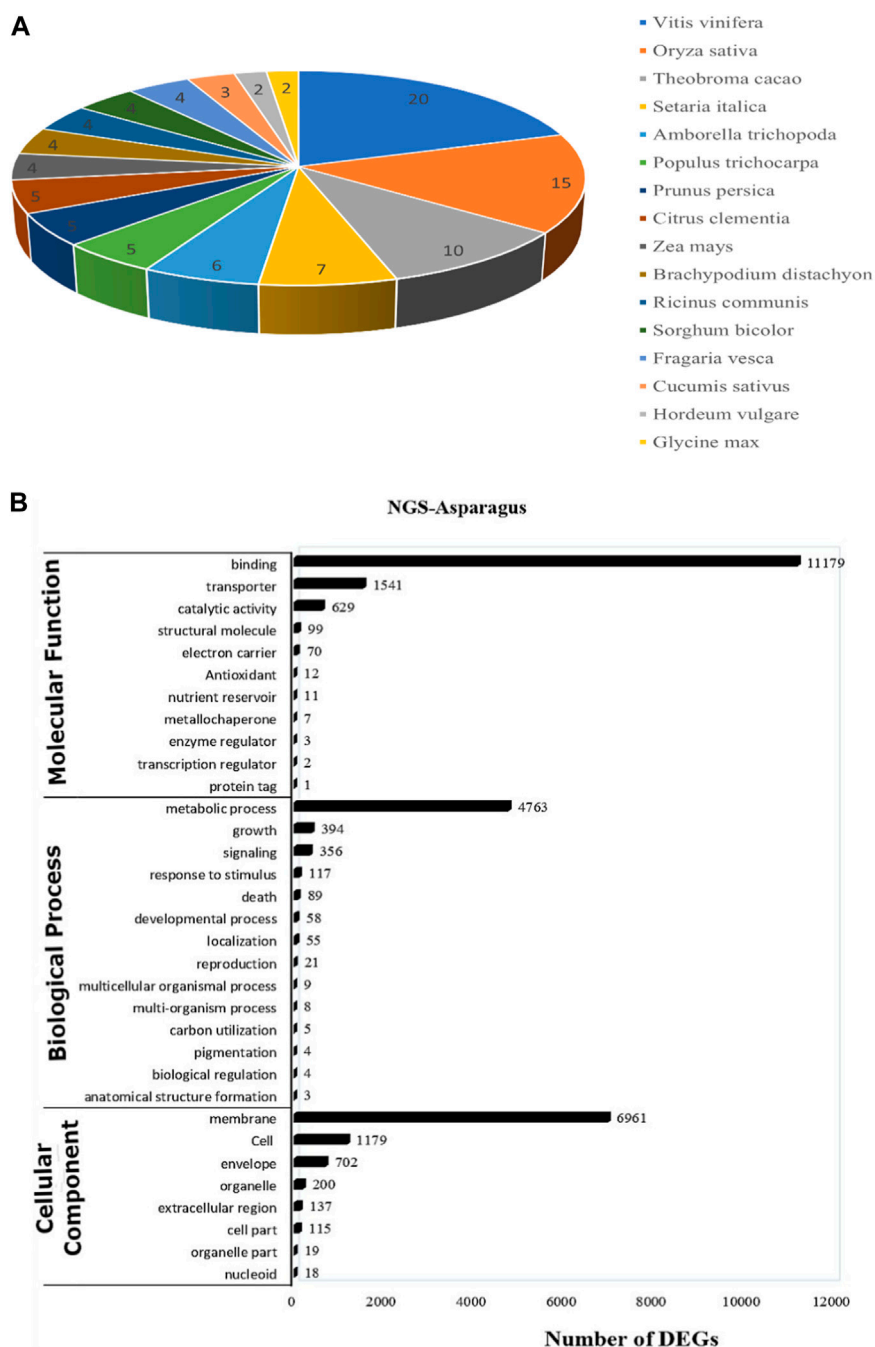


FIGURE 4

(A) The top blast hit species distribution of NGS-Asparagus sample Figure 4B (B) Gene ontology functional enrichment analysis across the three domains (Molecular Function, Biological process and Cellular Function) from RNA seq. data of *A. racemosus* under methyl jasmonate stress treatment.

tuberous root of *A. racemosus* has been traditionally used for medicinal purposes to treat various health conditions. The major secondary metabolites in the roots of *A. racemosus* are flavonoids, alkaloids, and saponins, which attributes to their pharmacological activities. Because of flavonoids, the plant's roots exhibit anticancer, antioxidant, anti-inflammatory, and anti-viral activities. Despite its applications, the genes involved in flavonoid biosynthesis pathways in *A. racemosus* are largely uncharacterized. Therefore, *de novo* transcriptome sequencing and assembly were performed in the

present study to explore genes involved in flavonoid biosynthesis pathways.

A large number of high-quality reads, transcripts, and CDS were obtained from *de novo* transcriptome sequencing data of *A. racemosus*. Accordingly, the current finding obtained ~2.31 million high-quality reads by paired-end (PE) 2 × 150bp library on Illumina NextSeq in pooled leaf and root libraries, significantly lower than the 60.77 million clean reads reported in *pear calli* (Premathilake et al., 2020). According to Fullwood et al.

TABLE 3 SSR markers analysis from RNA seq data of *A. racemosus* under methyl jasmonate stress treatment.

Description	NGS-Asparagus sample
Total number of Sequence examined	45,647
Total size of examined sequences (bp)	76,398,686
Total number of identified SSRs	7,558
Number of SSRs containing sequences	953
Number of SSRs present in compound formation	452
Unit Size of SSR	Number of SSRs
Di-nucleotide	3,810
Tri-nucleotide	3,486
Tetra-nucleotide	234
Penta-nucleotide	28
Total	7,558

(2009), to improve the efficacy of next-generation sequencing (NGS) for genome function and variation analysis, paired-end sequencing is significantly beneficial because it increases read length, improves accuracy, and reduces ambiguity compared to single-tag sequencing.

The identified CDS were searched against the NCBI non-redundant (NR) protein database, resulting in the annotation of 42,396 CDS. Most blast hits were found to be against *V. vinifera* and *O. sativa* showing closer relation with *A. racemosus*. Conesa et al. (2005) highlighted utilizing the Blast2GO program in conducting GO mapping, which assigns functions for BLASTX annotated CDS. Blast2GO is a powerful tool that enables automatic and high throughput functional annotation, making it suitable for non-model species genome research. The process of GO mapping includes searching for gene names or symbols in the gene-product tables of the GO database using BLASTX result accession IDs.

Additionally, accession IDs are used to retrieve UniProt IDs from Protein Information Resources (PIR) databases such as RefSeq, SwissProt, GenPept, TrEMBL, PSD, and PDB. Finally, the DBXREF table of the GO database is searched to obtain information on the accession IDs. The current finding functionally assigned many CDS to 25,342 GO terms, i.e., among the three main domains, a high number of CDS were allocated to Molecular Function (19,873), followed by Biological Process (19,550) and Cellular Component (14,577). The present report mapped a small number of CDS involved in molecular function than 99,868 CDS assigned for molecular function reported by Upadhyay et al. (2014) during *de novo* transcriptome analysis of novel genes involved in steroidal saponin biosynthesis of *A. racemosus*. In the GO functional annotation, the present finding assigned many GO terms (25,342) to CDS involved in flavonoids biosynthesis pathway, as compared to 7728 GO terms reported involved saponin biosynthesis pathways in the roots, leaves, and fruits of *A. racemosus* (Srivastava et al., 2018), suggests that various types GO terms are involved the flavonoid biosynthesis pathway. Also, in a study conducted by Loke et al. (2017) “Phenylpropanoid biosynthesis” pathway was found to

TABLE 4 KEGG pathways classification of CDS from RNA seq. of *A. racemosus* under methyl jasmonate stress treatment.

Pathway category	No. of CDS
Metabolism	
Carbon metabolism	382
Carbohydrate metabolism	524
Energy metabolism	340
Lipid metabolism	302
Nucleotide metabolism	241
Amino acid metabolism	381
Metabolism of other amino acids	158
Glycan biosynthesis and metabolism	142
Metabolism of cofactors and vitamins	271
Metabolism of terpenoids and polyketides	142
Biosynthesis of other secondary metabolites	117
Xenobiotics biodegradation and metabolism	65
Genetic information processing	
Transcription	394
Translation	645
Folding, sorting and degradation	522
Environmental information processing	
Replication and Repair	260
Membrane transport	22
Signal transduction	532
Signaling molecules and interaction	4
Cellular process	
Transport and catabolism	323
Cell motility	71
Cell growth and death	257
Cell communication	107
Sensory system	12
Organismal system	
Environmental adaptation	139

be particularly abundant in the root transcriptome. The process splits off into several significant pathways, including the flavonoid biosynthetic pathway, which contributes to plant resistance (Treutter, 2006). The direct involvement of phenylpropanoids in plant stress responses to temperature, drought, UV radiation, and nutrient deficiency is well established (Korkina, 2007).

KEGG differs from artificial intelligence and machine learning methods as it utilizes the cognitive ability of humans to construct “models” of biological systems, particularly in the form of KEGG pathway maps. These maps are created manually by extracting knowledge from published literature. They can be utilized in

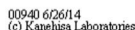
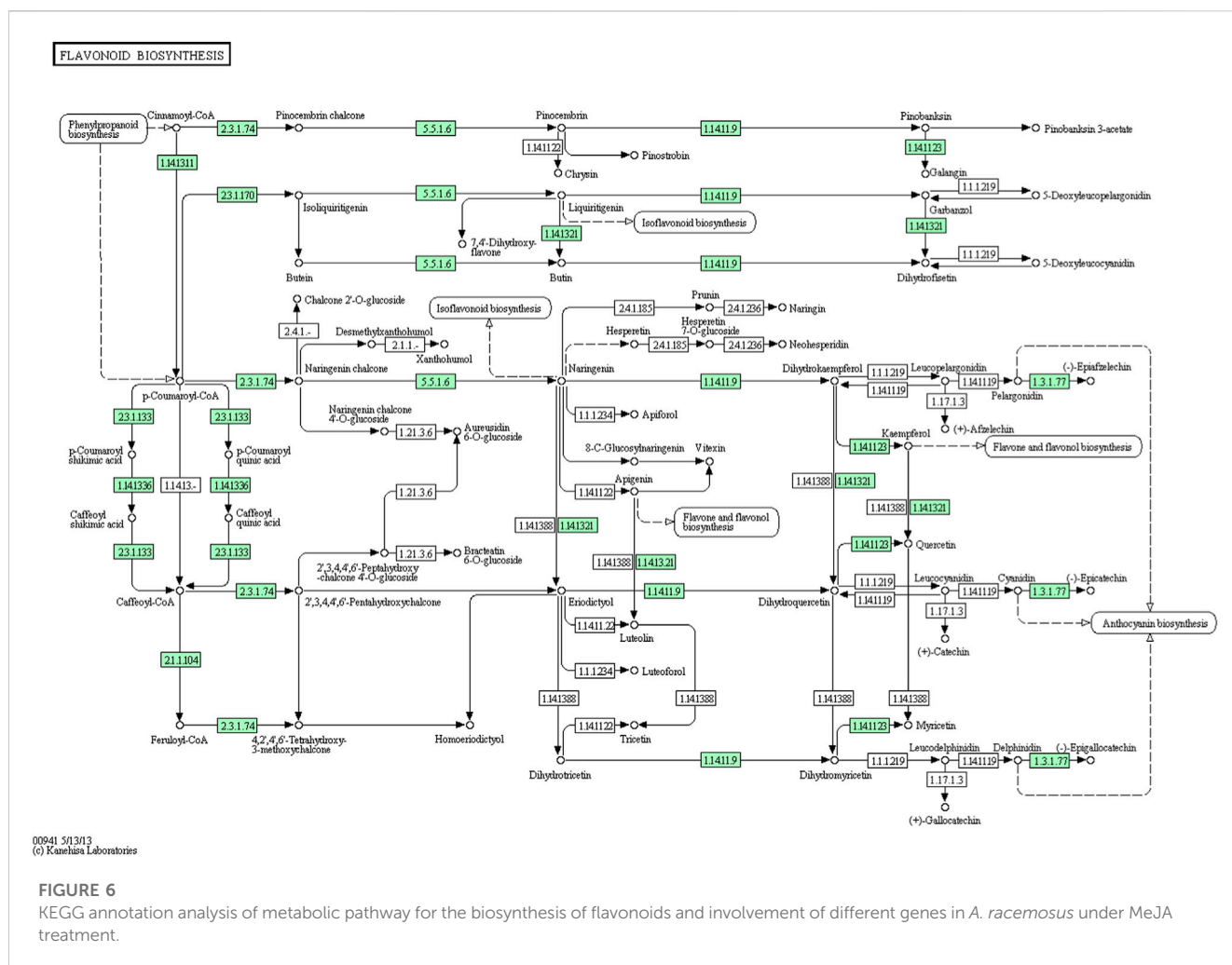


FIGURE 5
KEGG annotation analysis of metabolic pathway for the phenylpropanoid biosynthesis and involvement of different genes in *A. racemosus* under MeJA treatment.

According to Mazid et al. (2011), plants synthesize secondary metabolites to protect themselves from diverse biotic and abiotic stresses. The biosynthesis of these molecules can be induced using chemical elicitors, such as methyl jasmonate, which activate the biosynthetic genes through a complex signaling cascade. Shabani et al. (2009) highlighted that this process occurs at the transcriptional level, whereby methyl jasmonate transcriptionally activates the biosynthetic genes. The genes associated with flavonoid biosynthesis pathways were annotated and validated using qRT-PCR. A total of fourteen genes involved in flavonoid biosynthesis-associated were experimentally validated for their differential expression in pooled leaf and root samples of *A. racemosus* using

qRT-PCR. Most of the genes showed differential expressions in MeJA treated leaf sample than in the pooled leaf and root samples, suggesting flavonoid biosynthesis pathway genes are induced in response to the treatment of MeJA. The MeJA induced the expression of genes associated with the flavonoid biosynthesis pathway. The qRT-PCR revealed the differential accumulation genes in response to MeJA. The qRT-PCR resulted in a high-fold expression in the gene encoding an anthocyanidin-3-o-glycosyl transferase (FC = 15.029), followed by caffeoyl CoA methyltransferase (FC = 12.63) and quumoroil quinate monooxygenase (FC = 9.55) compared to the untreated plant. Anthocyanidin-3-o-glycosyl transferase (3 GT) is an enzyme that plays a critical role in the final step of anthocyanin biosynthesis by transferring a glucose, galactose, or arabinose molecule to the hydroxyl group of the anthocyanidin molecule, which enhances the stability and solubility of the anthocyanins (Rainha et al., 2020). Caffeoyl CoA methyltransferase (CCoAOMT) catalyzes the transfer of a methyl group from S-adenosyl-L-methionine (SAM) to the hydroxyl group of a phenolic compound, resulting in the formation of a methylated derivative. In flavonoid biosynthesis, CCoAOMT is involved in the conversion of caffeoyl-CoA to feruloyl-CoA, which is a key intermediate in the biosynthesis of a range of flavonoids (Hegde et al., 2021). Quumoroil quinate monooxygenase (CYP98A14) is a cytochrome P450 monooxygenase that catalyzes the hydroxylation of quinate to produce caffeoylquinate, a precursor



of chlorogenic acid and other phenolic compounds. In flavonoid biosynthesis, CYP98A14 is involved in the hydroxylation of p-coumaroylshikimate to form caffeoylshikimate, which is a key intermediate in the biosynthesis of a range of flavonoids, including anthocyanins, flavonols, and proanthocyanidins (Yang et al., 2022). This indicated that these three genes have significantly impacted the flavonoid biosynthesis pathway.

The qRT-PCR result exhibited low accumulations of genes encoding flavonol glycosyl transferase, flavanol synthase, and shikimate hydroxyl cinnamoyl transferase (cds_14257) in the MeJA-treated leaf sample. However, the gene encoding shikimate hydroxyl cinnamoyl transferase (cds_14258) was found to be downregulated in the MeJA-treated leaf sample. Upadhyay et al. (2014) found that the application of methyl jasmonate leads to an increase in plant accumulation by inducing gene transcripts related to secondary metabolites. Similarly, Premathilake et al. (2020) reported the upregulation of flavonoid biosynthesis pathway structural genes (PcCHS, PcCHI, PcF3H, PcDFR, PcANS, PcANR2a, and PcLARI) following treatment with MeJA. Du et al. (2021) studied the effect of methyl jasmonate on genes involved in flavonoid biosynthesis in pigeon pea plant and it was found that various genes including Caffeoyl CoA methyltransferase is induced by methyl jasmonate that supports current findings.

MISA microsatellite finder is a software application that is utilized for identifying microsatellites present in nucleotide sequences. It is not just limited to detecting flawless microsatellites, but is also able to identify perfect compound microsatellites, which comprise multiple instances of more than one simple sequence motif (Beier et al., 2017). In the present study, MISA microsatellite identification tool (MISA v1.0) from the pooled transcript resulted in 7,558 SSRs markers in the NGS-Asparagus sample, lower than the identified 18,107 in the leaf and 26,733 SSRs in the root of *A. racemosus* involved in steroidal apogonin biosynthesis in *A. racemosus* (Upadhyay et al., 2014). Furthermore, the di-nucleotide was exhibited as the most abundant, with a frequency of 50.41%, followed by trinucleotide, with a frequency distribution of 46.12%, which is not in agreement with the report of Upadhyay et al. (2014), which tri-nucleotide repeats were the most abundant SSR motif in leaf tissues followed by di-nucleotide. However, Srivastava et al. (2018) reported that mononucleotide (p1) SSR (50.93%, 20,800) was abundantly identified, followed by di-nucleotide (p2) (22.91%, 9356) and trinucleotide (p3) (17.37%, 7,096).

The transcriptome of *A. racemosus* was examined to identify transcripts encoding transcription factors (TFs). These TFs are part of various multi-gene families. They are essential in regulating the expression of individual or multiple genes by binding to

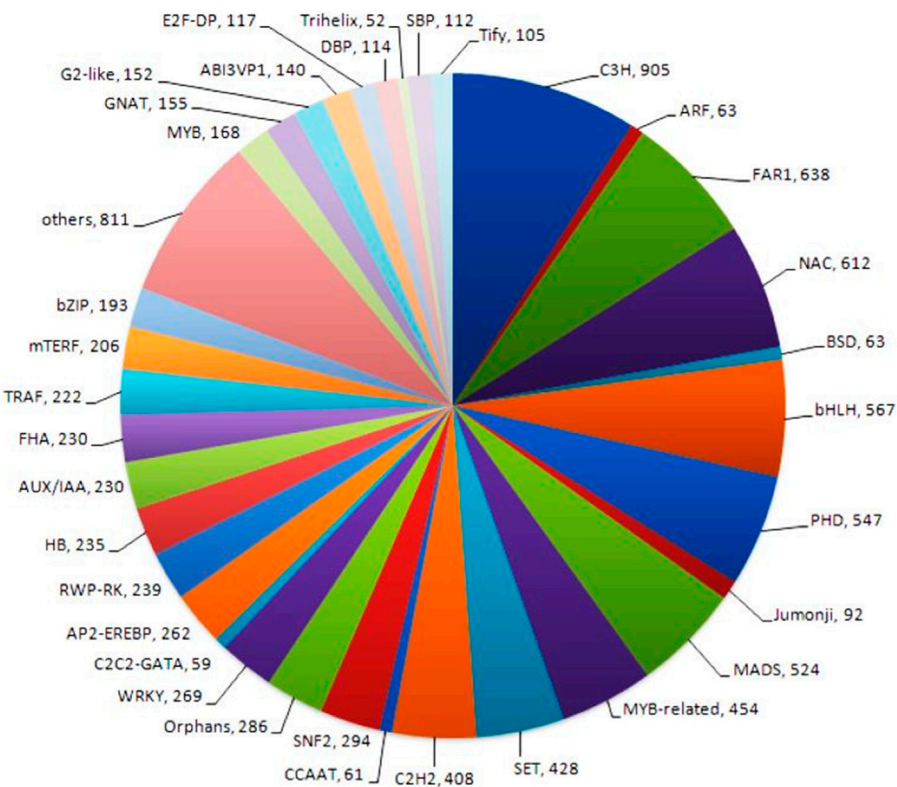


FIGURE 7 Transcription factor analysis generated 80 different transcription factor families using P2TF NGS-Asparagus sample Fold change.

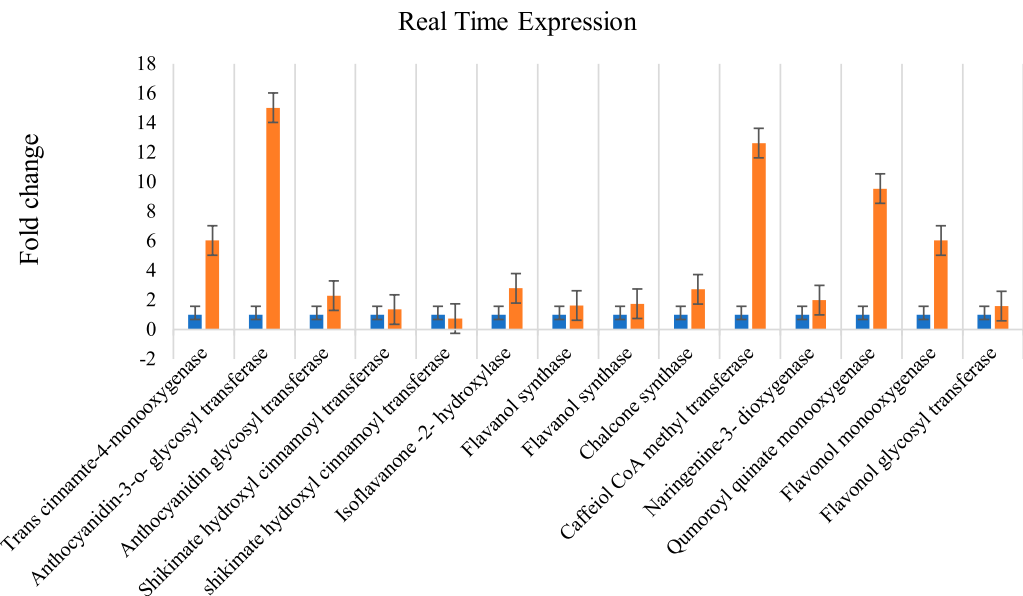


FIGURE 8 Methyl Jasmonate induced expression of genes involved in flavonoid biosynthesis pathways in *A. racemosus*. Orange colors represent the expression of genes associated with flavonoid biosynthesis pathways in MeJA-treated pooled leaf and root, while blue represents the control plant. Real Time Expression.

specific sequences in the promoter regions of target genes, known as cis-acting elements. (Broun et al., 2006; Pérez-Rodríguez et al., 2010). The current study revealed that the highest CDS (9.03%, 905) was assigned to Cysteine-3- Histidine (C3H), followed by a far-red-impaired response (FAR1) (6.37%, 638 CDS) and NAC transcription factors (6.11%, 612 CDS). The transcription factors belonging to the C3H-type zinc finger family are crucial in various aspects of plant growth and development and in their ability to adapt to biotic and abiotic stressors (Liu et al., 2020). FAR1 and NAC transcription factors families are more expressive in plants during stress conditions (Tran et al., 2007; Hou et al., 2023). However, Srivastava et al. (2018) reported the most significant members of transcription factors TIG (1,241, 10.55%), followed by FAR1 (925, 7.86%), C3H (785, 6.67%), MYB/MYB related (680, 5.78%). The present study assigned 6.37%, 638 CDS to C3H and 4.54%, 454 CDS, which is associated with Srivastava et al. (2018) reports that C3H and MYB/MYB related assigned (785, 6.67%) and (680, 5.78%) in *A. racemosus* transcriptome, respectively. The transcriptome analysis of *A. racemosus* further revealed the presence of several transcription factors (TF) families, including AP2-EREBP, bHLH, C2C2, MYB/MYB related, and WRKY, which have been previously reported to regulate the biosynthesis of secondary metabolites in plants (Kato et al., 2007; Suttipanta et al., 2011; Zhang et al., 2011). The present finding could give an insight into the molecular understanding of the gene involved in the flavonoid biosynthesis pathway and helps for further secondary metabolic improvement in the *A. racemosus*.

5 Conclusion

The EST data obtained from *de novo* transcriptome sequencing provides the foundation for functional genomics research in *A. racemosus*. The annotated transcripts provide valuable information for studying the biochemical, cellular, and molecular aspects of the *A. racemosus* transcriptome. This knowledge can aid in metabolic engineering and elucidate the regulation of secondary metabolic biosynthetic pathways in plants. The gene encoding anthocyanidin-3-o-glycosyl transferase, caffeoyl CoA methyltransferase, and quumoyl quinate monooxygenase have a significant role in the biosynthesis of the flavonoid pathway. Methyl jasmonate has been found to be an effective inducer of flavonoid biosynthesis transcripts in plants, resulting in the upregulation of genes encoding enzymes involved in flavonoid biosynthesis. The expression of the gene encoding Shikimate hydroxyl cinnamoyl transferase (cds_14258) was inhibited by the treatment of MeJA and downregulated in the treated sample. Understanding the regulation of flavonoid biosynthesis and the key enzymes involved in this process can aid in developing new strategies for plant metabolic engineering and producing high-value flavonoid compounds. The limitation of current study is that it only provides information on a part of the whole picture as only a small number of potential genes involved in flavonoid biosynthesis were chosen for research. Although there are various other candidate genes which are involved in flavonoid biosynthesis identified by *de novo* sequencing and are worthy of further investigation. Consequently, it makes it hard to draw precise

conclusions. However, the dataset will provide an important resource foundation for future genetic or genomes studies on *Asparagus* species and will help to give better insight into the mechanism of biosynthesis of flavonoids in *A. racemosus*. This information can significantly help genetically improve and identify germplasm with high alkaloids, flavonoids and saponins.

Data availability statement

The original contributions presented in the study are publicly available. This data can be found here: <https://www.ncbi.nlm.nih.gov/bioproject/>. Accession number: PRJNA977204.

Author contributions

VC and CM conceived the current research work. CM and SD collected sample plant material and performed RNA extraction followed by RT-PCR. CM, TR, RK, AK and RY analyzed the data obtained. CM, SD and TR prepared the original draft, which was further reviewed and edited by VC, RY, NS and RK. The whole study was done under the supervision of VC. All authors approved the final draft of the manuscript after they had reviewed the findings.

Acknowledgments

CM is grateful to the Department of Bio and Nano Technology, Guru Jambheshwar University of Science and Technology, for providing laboratory facilities and also to Xcelris Genomics, Ahmedabad, Gujarat, India, for *de novo* sequencing of *A. racemosus*.

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/fgene.2023.1236517/full#supplementary-material>

References

- Aggarwal, H., Gyanprakash, R. A., and Chhokar, V. (2013). Evaluation of root extracts of *Asparagus racemosus* for antibacterial activity. *Am. J. Drug Discov. Dev.* 3 (2), 113–119. doi:10.3923/ajdd.2013.113.119
- Alok, S., Jain, S. K., Verma, A., Kumar, M., Mahor, A., and Sabharwal, M. (2013). Plant profile, phytochemistry and pharmacology of *Asparagus racemosus* (Shatavari): A review. *Asian Pac. J. Trop. Dis.* 3 (3), 242–251. doi:10.1016/S2222-1808(13)60049-3
- Beier, S., Thiel, T., Münch, T., Scholz, U., and Mascher, M. (2017). MISA-Web: A web server for microsatellite prediction. *Bioinformatics* 33 (16), 2583–2585. doi:10.1093/bioinformatics/btx198
- Bopana, N., and Saxena, S. (2008). *In vitro* propagation of a high value medicinal plant: *asparagus racemosus* willd. *Vitro Cell. Dev. Biol. Plant* 44, 525–532. doi:10.1007/s11627-008-9137-y
- Broun, P., Liu, Y., Queen, E., Schwarz, Y., Abenes, M. L., and Leibman, M. (2006). Importance of transcription factors in the regulation of plant secondary metabolism and their relevance to the control of terpenoid accumulation. *Phytochem. Rev.* 5, 27–38. doi:10.1007/s11101-006-9000-x
- Chawla, A., Chawla, P., Mangalesh, R. R., and Roy, R. (2011). *Asparagus racemosus* (willd): biological activities and its active principles. *Indo-Global J. Pharm. Sci.* 2, 113–120. doi:10.35652/igjps.2011.11
- Choudhri, P., Rani, M., Sangwan, R. S., Kumar, R., Kumar, A., and Chhokar, V. (2018). De novo sequencing, assembly and characterisation of *Aloe vera* transcriptome and analysis of expression profiles of genes related to saponin and anthraquinone metabolism. *BMC Genom* 19, 427–521. doi:10.1186/s12864-018-4819-2
- Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., and Robles, M. (2005). Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21 (18), 3674–3676. doi:10.1093/bioinformatics/bti610
- Du, T., Fan, Y., Cao, H., Song, Z., Dong, B., Liu, T., et al. (2021). Transcriptome analysis revealed key genes involved in flavonoid metabolism in response to jasmonic acid in pigeon pea (*Cajanus cajan* (L) Millsp). *Plant Physiol. biochem.* 168, 410–422. doi:10.1016/j.plaphy.2021.10.022
- Ferrer, J. L., Austin, M. B., Stewart, C., Jr, and Noel, J. P. (2008). Structure and function of enzymes involved in the biosynthesis of phenylpropanoids. *Plant Physiol. Biochem.* 46 (3), 356–370. doi:10.1016/j.plaphy.2007.12.009
- Fullwood, M. J., Wei, C. L., Liu, E. T., and Ruan, Y. (2009). Next-generation DNA sequencing of paired-end tags (PET) for transcriptome and genome analyses. *Genome Res.* 19 (4), 521–532. doi:10.1101/gr.074906.107
- Gupta, M., and Shaw, B. (2011). A double-blind randomized clinical trial for evaluation of galactogogue activity of *Asparagus racemosus* willd. *Iran. J. Pharm. IJPR* 10 (1), 167–172.
- Hayes, P. Y., Jahidin, A. H., Lehmann, R., Penman, K., Kitching, W., and De Voss, J. J. (2008). Steroidal saponins from the roots of *Asparagus racemosus*. *Phytochemistry* 69 (3), 796–804. doi:10.1016/j.phytochem.2007.09.001
- Hegde, N., Joshi, S., Soni, N., and Kushalappa, A. C. (2021). The caffeoyl-CoA O-methyltransferase gene SNP replacement in Russet Burbank potato variety enhances late blight resistance through cell wall reinforcement. *Plant Cell Rep.* 40, 237–254. doi:10.1007/s00299-020-02629-6
- Hossain, M. I., Sharmin, F. A., Akhter, S., Bhuiyan, M. A., and Shahriar, M. (2012). Investigation of cytotoxicity and *in-vitro* antioxidant activity of *Asparagus racemosus* root extract. *Int. Curr. Pharm. J.* 1 (9), 250–257. doi:10.3329/icpj.v1i9.11615
- Hou, R. M., Yang, L. Z., Wuyun, T., Chen, S. Y., and Zhang, L. (2023). Genes related to osmoregulation and antioxidation play important roles in the response of *Trollius chinensis* seedlings to saline-alkali stress. *Front. Plant Sci.* 14, 65. doi:10.3389/fpls.2023.1080504
- Jian, B., Liu, B., Bi, Y., Hou, W., Wu, C., and Han, T. (2008). Validation of internal control for gene expression study in soybean by quantitative real-time PCR. *BMC Mol. Biol.* 9 (1), 59–14. doi:10.1186/1471-2199-9-59
- Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y., and Hattori, M. (2004). The KEGG resource for deciphering the genome. *Nucleic Acids Res.* 32, D277–D280. doi:10.1093/nar/gkh063
- Kanehisa, M., Sato, Y., and Kawashima, M. (2022). KEGG mapping tools for uncovering hidden features in biological data. *Protein Sci.* 31 (1), 47–53. doi:10.1002/pro.4172
- Kato, N., Dubouzet, E., Kokabu, Y., Yoshida, S., Taniguchi, Y., Dubouzet, J. G., et al. (2007). Identification of a WRKY protein as a transcriptional regulator of benzyloisoquinoline alkaloid biosynthesis in *Coptis japonica*. *Plant Cell Physiol.* 48 (1), 8–18. doi:10.1093/pcp/pcl041
- Korkina, L. G. (2007). Phenylpropanoids as naturally occurring antioxidants: from plant defense to human health. *Cell. Mol. Biol.* 53 (1), 15–25.
- Ku, K. M., and Juvik, J. A. (2013). Environmental stress and methyl jasmonate-mediated changes in flavonoid concentrations and antioxidant activity in broccoli florets and kale leaf tissues. *Hortscience* 48 (8), 996–1002. doi:10.21273/hortsci.48.8.996
- Kumar Kar, D., and Sen, S. (1985). Propagation of *Asparagus racemosus* through tissue culture. *Plant Cell, Tissue Organ Cul* 5, 89–95. doi:10.1007/bf00033574
- Kumar, M., Kumar Naik, P., Patlan, S., and Chhokar, V. (2016). Assessment of genetic variation among *Asparagus racemosus* genotypes using molecular and biochemical markers. *J. Med. Plants Stud.* 6, 117–123.
- Kumar, M., Sarla Yadav, O. P., and Chhokar, V. (2011). Estimation of saponin in *Asparagus (Asparagus racemosus)* roots by colorimetric method. *Ann. Biol.* 27 (2), 115–120.
- Li, H., Dong, Y., Yang, J., Liu, X., Wang, Y., Yao, N., et al. (2012). De novo transcriptome of safflower and the identification of putative genes for oleosin and the biosynthesis of flavonoids. *PLoS one* 7 (2), e30987. doi:10.1371/journal.pone.0030987
- Liu, C., Xu, X., Kan, J., ming Cheng, Z., Chang, Y., Lin, J., et al. (2020). Genome-wide analysis of the C3H zinc finger family reveals its functions in salt stress responses of *Pyrus betulaeifolia*. *PeerJ* 8, e9328. doi:10.7717/peerj.9328
- Liu, J. P., Xia, Z. Q., Tian, X. Y., and Li, Y. J. (2015). Transcriptome sequencing and analysis of rubber tree (*Hevea brasiliensis* Muell) to discover putative genes associated with tapping panel dryness (TPD). *BMC genom* 16, 398–413. doi:10.1186/s12864-015-1562-9
- Loke, K. K., Rahnamaie-Tajadod, R., Yeoh, C. C., Goh, H. H., Mohamed-Hussein, Z. A., Zainal, Z., et al. (2017). Transcriptome analysis of *Polygonum minus* reveals candidate genes involved in important secondary metabolic pathways of phenylpropanoids and flavonoids. *PeerJ* 5, e2938. doi:10.7717/peerj.2938
- Mazid, M., Khan, T. A., and Mohammad, F. (2011). Role of secondary metabolites in defense mechanisms of plants. *Biol. Med.* 3 (2), 232–249.
- Mitra, S. K., Prakash, N. S., and Sundaram, R. (2012). Shatavarins (containing Shatavarin IV) with anticancer activity from the roots of *Asparagus racemosus*. *Indian J. Pharmacol.* 44 (6), 732–736. doi:10.4103/0253-7613.103273
- Mizrachi, E., Hefer, C. A., Ranik, M., Joubert, F., and Myburg, A. A. (2010). De novo assembled expressed gene catalog of a fast-growing *Eucalyptus* tree produced by Illumina mRNA-Seq. *BMC genom* 11 (1), 681–712. doi:10.1186/1471-2164-11-681
- Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C., and Kanehisa, M. (2007). Kaas: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* 35 (2), W182–W185. doi:10.1093/nar/gkm321
- Pérez-Rodríguez, P., Riano-Pachon, D. M., Corréa, L. G. G., Rensing, S. A., Kersten, B., and Mueller-Roeber, B. (2010). PlnTFDB: updated content and new features of the plant transcription factor database. *Nucleic Acids Res.* 38 (1), D822–D827. doi:10.1093/nar/gkp805
- Prakash, A., and Singh, K. K. (2006). Useful medicinal plants for general tonic, sexual vigour, strength, vitality and venereal diseases among the tribals of Uttar Pradesh and Uttaranchal. *J. Econ. Taxon. Bot.* 30, 326–331. Scientific Publishers, (India).
- Premathilake, A. T., Ni, J., Shen, J., Bai, S., and Teng, Y. (2020). Transcriptome analysis provides new insights into the transcriptional regulation of methyl jasmonate-induced flavonoid biosynthesis in pear calli. *BMC plant Biol.* 20 (1), 388–414. doi:10.1186/s12870-020-02606-x
- Rainha, J., Gomes, D., Rodrigues, L. R., and Rodrigues, J. L. (2020). Synthetic biology approaches to engineer *Saccharomyces cerevisiae* towards the industrial production of valuable polyphenolic compounds. *Life* 10 (5), 56. doi:10.3390/life10050056
- Rani, M., Jaglan, S., Beniwal, V., and Chhokar, V. (2022). Bioactive saponin profiling of endophytic fungi from *Asparagus racemosus*. *Nat. Prod. Res.* 1–7. doi:10.1080/14786419.2022.2156997
- Schmittgen, T. D., and Livak, K. J. (2008). Analyzing real-time PCR data by the comparative CT method. *Nat. Protoc.* 3 (6), 1101–1108. doi:10.1038/nprot.2008.73
- Shabani, L., Ehsanpour, A. A., Asghari, G., and Emami, J. (2009). Glycyrrhizin production by *in vitro* cultured *Glycyrrhiza glabra* elicited by methyl jasmonate and salicylic acid. *Russ. J. Plant Physiol.* 56, 621–626. doi:10.1134/S1021443709050069
- Sharma, K., and Bhatnagar, M. (2010). *Asparagus racemosus* (Shatavari): A versatile female tonic. *health* 3 (4), 5–6.
- Srivastava, P. L., Shukla, A., and Kalunke, R. M. (2018). Comprehensive metabolic and transcriptomic profiling of various tissues provide insights for saponin biosynthesis in the medicinally important *Asparagus racemosus*. *Sci. Rep.* 8 (1), 9098–9113. doi:10.1038/s41598-018-27440-y
- Sun, C., Li, Y., Wu, Q., Luo, H., Sun, Y., Song, J., et al. (2010). De novo sequencing and analysis of the American ginseng root transcriptome using a GS FLX Titanium platform to discover putative genes involved in ginsenoside biosynthesis. *BMC genom* 11, 262–312. doi:10.1186/1471-2164-11-262
- Suttipanta, N., Pattanaik, S., Kulshrestha, M., Patra, B., Singh, S. K., and Yuan, L. (2011). The transcription factor CrWRKY1 positively regulates the terpenoid indole alkaloid biosynthesis in *Catharanthus roseus*. *Plant physiol.* 157 (4), 2081–2093. doi:10.1104/pp.111.181834
- Taeponsorot, L., and Rattana, S. (2018). Antioxidant activities of ethanolic and aqueous extracts of *Asparagus racemosus* Roots. *Pharmacogon J.* 10 (6), 1129–1132. doi:10.5530/pj.2018.6.192

- Tran, L. S. P., Nakashima, K., Sakuma, Y., Osakabe, Y., Qin, F., Simpson, S. D., et al. (2007). Co-expression of the stress-inducible zinc finger homeodomain ZFHD1 and NAC transcription factors enhances expression of the ERD1 gene in Arabidopsis. *Plant J* 49 (1), 46–63. doi:10.1111/j.1365-313X.2006.02932.x
- Treutter, D. (2006). Significance of flavonoids in plant resistance: A review. *Environ. Chem. Lett.* 4 (3), 147–157. doi:10.1007/s10311-006-0068-8
- Upadhyay, S., Phukan, U. J., Mishra, S., and Shukla, R. K. (2014). De novo leaf and root transcriptome analysis identified novel genes involved in steroidal sapogenin biosynthesis in *Asparagus racemosus*. *BMC genom* 15 (1), 746–813. doi:10.1186/1471-2164-15-746
- Ververidis, F., Trantas, E., Douglas, C., Vollmer, G., Kretschmar, G., and Panopoulos, N. (2007). Biotechnology of flavonoids and other phenylpropanoid derived natural products. Part I: chemical diversity, impacts on plant biology and human health. *Biotechnol. J. Healthc. Nutr. Technol.* 2 (10), 1214–1234. doi:10.1002/biot.200700084
- Yang, J., Li, H., Ma, R., Chang, Y., Qin, X., Xu, J., et al. (2022). Genome-wide transcriptome analysis and characterization of the cytochrome P450 flavonoid biosynthesis genes in pigeon pea (*Cajanus cajan*). *Planta* 255 (6), 120. doi:10.1007/s00425-022-03896-1
- Zhang, H., Hedhili, S., Montiel, G., Zhang, Y., Chatel, G., Pré, M., et al. (2011). The basic helix-loop-helix transcription factor CrMYC2 controls the jasmonate-responsive expression of the ORCA genes that regulate alkaloid biosynthesis in *Catharanthus roseus*. *Plant J* 67 (1), 61–71. doi:10.1111/j.1365-313X.2011.04575.x



OPEN ACCESS

EDITED BY

Krishnanand P. Kulkarni,
Delaware State University, United States

REVIEWED BY

Vignesh Muthusamy,
Indian Agricultural Research Institute
(ICAR), India
Willy Bayuardi Suwarno,
IPB University, Indonesia

*CORRESPONDENCE

Manje Gowda,
✉ m.gowda@cgjar.org

[†]These authors have contributed equally
to this work

RECEIVED 24 July 2023

ACCEPTED 12 October 2023

PUBLISHED 26 October 2023


CITATION

Kimutai C, Ndlovu N, Chaikam V,
Ertiro BT, Das B, Beyene Y, Kiplagat O,
Spillane C, Prasanna BM and Gowda M
(2023), Discovery of genomic regions
associated with grain yield and
agronomic traits in Bi-parental
populations of maize (*Zea mays*. L) Under
optimum and low nitrogen conditions.
Front. Genet. 14:1266402.
doi: 10.3389/fgene.2023.1266402

COPYRIGHT

© 2023 Kimutai, Ndlovu, Chaikam, Ertiro,
Das, Beyene, Kiplagat, Spillane, Prasanna
and Gowda. This is an open-access article
distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). 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.

Discovery of genomic regions associated with grain yield and agronomic traits in Bi-parental populations of maize (*Zea mays*. L) Under optimum and low nitrogen conditions

Collins Kimutai^{1,2†}, Noel Ndlovu ^{3†}, Vijay Chaikam ²,
Berhanu Tadesse Ertiro², Biswanath Das², Yoseph Beyene²,
Oliver Kiplagat¹, Charles Spillane ³, Boddupalli M. Prasanna ²
and Manje Gowda ^{2*}

¹Seed, Crop and Horticultural Sciences, School of Agriculture and Biotechnology, University of Eldoret, Eldoret, Kenya, ²International Maize and Wheat Improvement Center (CIMMYT), Nairobi, Kenya,

³Agriculture and Bioeconomy Research Centre, Ryan Institute, University of Galway, Galway, Ireland

Low soil nitrogen levels, compounded by the high costs associated with nitrogen supplementation through fertilizers, significantly contribute to food insecurity, malnutrition, and rural poverty in maize-dependent smallholder communities of sub-Saharan Africa (SSA). The discovery of genomic regions associated with low nitrogen tolerance in maize can enhance selection efficiency and facilitate the development of improved varieties. To elucidate the genetic architecture of grain yield (GY) and its associated traits (anthesis-silking interval (ASI), anthesis date (AD), plant height (PH), ear position (EPO), and ear height (EH)) under different soil nitrogen regimes, four F₃ maize populations were evaluated in Kenya and Zimbabwe. GY and all the traits evaluated showed significant genotypic variance and moderate heritability under both optimum and low nitrogen stress conditions. A total of 91 quantitative trait loci (QTL) related to GY (11) and other secondary traits (AD (26), PH (19), EH (24), EPO (7) and ASI (4)) were detected. Under low soil nitrogen conditions, PH and ASI had the highest number of QTLs. Furthermore, some common QTLs were identified between secondary traits under both nitrogen regimes. These QTLs are of significant value for further validation and possible rapid introgression into maize populations using marker-assisted selection. Identification of many QTL with minor effects indicates genomic selection (GS) is more appropriate for their improvement. Genomic prediction within each population revealed low to moderately high accuracy under optimum and low soil N stress management. However, the accuracies were higher for GY, PH and EH under optimum compared to low soil N stress. Our findings indicate that genetic gain can be improved in maize breeding for low N stress tolerance by using GS.

KEYWORDS

grain yield, low soil nitrogen, maize, sub-Saharan Africa, quantitative trait loci (QTL)

1 Introduction

Maize (*Zea mays* L.), a staple crop in sub-Saharan Africa (SSA), is projected to increase in demand (Ekpa et al., 2018), necessitating improved breeding systems to maximize productivity. Over the years, maize productivity in SSA has been limited by the adverse effects of parasitic weeds (e.g., *Striga* sp. (Kanampiu et al., 2018; Yacoubou et al., 2021)), insect infestations (e.g., fall armyworm (Baudron et al., 2019; De Groote et al., 2020)), and disease incidences (e.g., Maize lethal necrosis (Boddupalli et al., 2020)). Most notably, inadequate soil fertility (particularly low nitrogen) is a primary cause of low maize yields in the region's smallholder farming communities. Modelling studies have established that climate change has the potential to worsen these current maize production constraints (Tesfaye et al., 2015), especially under low soil nitrogen conditions (Falconnier et al., 2020). In this regard, maize breeding holds a significant potential to provide sustainable solutions to the prevailing and projected biotic and abiotic constraints (Eriksson et al., 2018; Ndlovu, 2020).

As nitrogen use efficiency (NUE) is critical for sustainable productivity, there is major interest in developing varieties that perform better in low soil nitrogen stress conditions (Ribaut et al., 2007). However, to harness plant breeding for such stress conditions, an understanding of the performance of maize genotypes under low nitrogen conditions is required. Assessing the performance of maize genotypes under different nitrogen regimes is critical for identifying promising parental lines. Phenotypic evaluations can reveal useful diversity in maize germplasm (Shitta et al., 2021; Chen et al., 2023a). However, morphological characterizations of genotypes are influenced by both the genotype (G) and environment (E), as well as G \times E interaction, and hence may not accurately reflect the genotypic effect of nitrogen utilization *per se* (de Carvalho et al., 2012). In addition, large genotype \times season and genotype \times location interactions can stymie progress in selecting for low nitrogen tolerance (Ribaut et al., 2007). In this context, understanding the genetic architecture of grain yield (GY) and other secondary traits that are associated with low soil nitrogen can speed up genetic improvement (Ertiro et al., 2020c) for low nitrogen tolerance in tropical maize.

Maize GY under low soil nitrogen is a complex trait governed by multiple genes. In most cases, maize breeding for low nitrogen focuses on the anthesis-silking interval (ASI), anthesis date (AD), plant height (PH), ear position (EPO), and ear height (EH) for indirect selection and GY for direct selection (Worku et al., 2007; Worku et al., 2008; Worku et al., 2012; Das et al., 2019; Ertiro et al., 2020b; Ertiro et al., 2022; Ndlovu et al., 2022). However, morphological characterizations are time-consuming, cost-ineffective, and lack accuracy. The complex nature of breeding maize under low nitrogen conditions has necessitated the use of quantitative trait loci (QTL) based approaches which have been shown to contribute significantly to the understanding of the genetic basis of crop performance and stability under nitrogen-stressed conditions (Semagn et al., 2015). The discovery and characterization of QTL can assist breeders in using genomic regions linked with complex trait expression and deciphering their genetic contribution at the target loci (Ertiro et al., 2020c). In addition, through QTL analysis, the biological mechanisms responsible for phenotypic expression can be pursued.

Furthermore, exploiting molecular markers in breeding has allowed the mapping of a subset of markers associated with one or more QTL that contribute to regulating the expression of complex traits like GY. Such molecular markers can form the concrete basis for the use of other genomic approaches such as genomic selection in the desired population (Würschum, 2012). Unfortunately, little is known about the genetic basis of GY and associated traits under low nitrogen conditions, and major effect QTLs are yet to be reported (Ertiro et al., 2020c).

Genomic selection (GS) is a tool increasingly used to reduce the breeding cycle length and increase genetic gain in both crops and livestock (Meuwissen et al., 2001). The success of GS in the dairy industry (where the breeding cycle of the cattle can be reduced from 7 years to 18 months) has helped to achieve twice the genetic gain for key traits (García-Ruiz et al., 2016). Such successes have motivated plant breeders to apply GS in crop improvement for reducing the breeding cycle and increase genetic gain. Currently, GS is applied to improve a range of different traits in maize (Beyene et al., 2015; Ertiro et al., 2020a; Kibe et al., 2020; Gowda et al., 2021; Ndlovu et al., 2022), wheat (Dreisigacker et al., 2021; Bonnett et al., 2022; Juliana et al., 2022; Ficht et al., 2023), and other crops (Hu et al., 2022; Qin et al., 2022; Chen et al., 2023b). GS has been successfully integrated into maize breeding programs for improving GY under optimum and drought conditions (Beyene et al., 2015; Zhang et al., 2017b; Vivek et al., 2017; Beyene et al., 2019; Wang et al., 2020; Beyene et al., 2021). Linkage mapping enables the detection of QTL for the target trait by using different bi-parental populations, whereas GS enables the selection of superior individuals by considering the effects of multiple genes controlling a target trait (Crossa et al., 2017; Yuan et al., 2019). Combining linkage mapping results with GS will accelerate the breeding efficiency for GY and other complex traits under low N stress conditions. Therefore, the full potential of GS needs to be assessed in biparental and/or practical breeding populations for low soil N tolerance.

Studies conducted over the last 20 years have identified QTL that generally explained a significant portion of the phenotypic variance, and therefore gave rise to an optimistic assessment of the prospects of marker-assisted selection (Semagn et al., 2010). According to Coque and Gallais (2006), the detected marker QTL associations in maize revealed the consistency of the involvement of some traits, such as root architecture and glutamine synthase activity, which would be of major importance for GY setting under both optimum and low nitrogen conditions. The discovery of genomic regions associated with GY and other agronomic traits in maize under low soil nitrogen conditions is of paramount importance. Hence, this study sought to evaluate four F3 populations to 1) estimate the phenotypic effect and heritability for GY and other agronomic traits under optimum and low nitrogen management 2) identify the genomic regions associated with these traits, and 3) assessing the usefulness of GS in improving GY and other agronomic traits under optimum and low soil N conditions. The findings of this study can provide genetic resources that can be used in scaling the application of MAS for enhancing maize GY in nitrogen-starved soils in SSA. Furthermore, the identification of QTL can hasten maize breeding cycles and facilitate the release of nitrogen-tolerant or NUE varieties for resource-constrained smallholder farmer communities in the region.

TABLE 1 Details of the populations used in this study and number of locations evaluated under optimum and low N management.

Population	Pedigree	Tester	Population size	No of locations	
				Opt	Low N
POPULATION 1	CML494×CML550	CML495	357	3	4
POPULATION 2	CKL05017× CML536	CML444× CML395	276	3	3
POPULATION 3	CML550× CML507	CML442× CML312	315	2	1
POPULATION 4	VL081452× VL058589	CML444× CML395	158	1	2

*Opt, Optimum; Low N–Low soil N management.

2 Materials and methods

2.1 Plant materials

Four F3 tropical maize populations were evaluated in Kenya and Zimbabwe. The populations were developed by the Global Maize Program of the International Maize and Wheat Improvement Center (CIMMYT). The specific details of the tested populations are presented in Table 1. These lines are adapted to mid-altitude regions (1,000–1500 M Above sea level) of SSA. They are bred on a good-by-good basis and adapted to stress conditions. Populations are test crossed with an appropriate tester from the opposite heterotic group for phenotypic evaluations. The CIMMYT lines utilized for the project were derived from breeding programs targeting tolerance to low soil nitrogen; hence the best choice for QTL mapping with most of the lines included in two association mapping panel (AMP) constituted under IMAS (improved maize for Africa soils) and DTMA (Drought Tolerant Maize for Africa) projects (Semagn et al., 2012; Ertiro et al., 2020a). The four populations formed a set of multiple bi-parental populations used in the current study.

2.2 Field trial

All four populations were evaluated for response to low soil nitrogen conditions at one to four different locations; Kiboko (Longitude 37°E, Latitude 2°S, 975 M A.S.L), Embu (Longitude 37°E, Latitude 3°S, 1560 M A.S.L), and Harare in Zimbabwe (Longitude 31° E, Latitude 17°S, 1490 M A.S.L). The lines were evaluated on an alpha lattice incomplete block design under two nitrogen levels. The two nitrogen treatments were low N (N-depleted field/plot with no application of N fertilizer) and at normal farmer practice conditions (optimum N 200 kg/ha). The low-N sites were prepared by depleting N by growing sorghum at high density with no N fertilizer sources added for four cropping cycles. The depletion crop was uprooted at near maturity and removed from the low-N field to prevent the incorporation of crop residues into the soil. A soil nitrate concentration ranging from 7.5 to 15 parts per million (ppm) is indicative of soil nitrogen deficiency (Zaman-Allah et al., 2018). Low N sites with around 10 ppm of nitrate level provide good experimental conditions for detecting useful genetic variation. The sites used in the present study showed soil nitrate levels ranging from 0.10 to 0.12 ppm in stress experimental sites, while in the optimum sites, the levels were greater

than 0.25 ppm. Single row plots measuring 5 m long at 0.75 m row spacing with two seeds per hill were sown. After 3 weeks of planting, plants were thinned to one plant per hill to obtain a final population density of 53,333 plants per hectare. All entries were planted on the same day in conventionally tilled plots and maintained under rain-fed or irrigated conditions.

2.3 Data collection

2.3.1 Phenotyping of important agronomic traits

Ten plants in the middle of the row were selected for each genotype for phenotypic evaluation. Phenotypic components measured and analyzed are plant height (PH in centimeters), anthesis date (AD, days), anthesis silking interval (ASI, days), ear height (EH, in centimeters), ear position (EPO, ratio of EH/PH), number of ears per plant (EPP, total number of ears per plot divided by number of plants per plot) and grain yield (GY in t/ha). Mature ears were harvested, manually bagged, air-dried, and shelled on an electric shelling device. The total GY of each plot was weighed on a balance and converted to GY into t/ha.

Best linear unbiased predictors (BLUPs) were calculated with mixed model where genotypes and other factors were treated as random except replications and best linear unbiased estimators (BLUEs) were calculated where genotype entries and replications are treated as fixed effects and the rest of the terms as random. Estimating broad-sense heritability, all the terms were considered random. Broad sense heritability was estimated by the formula:

$$h^2 = \sigma_G^2 / (\sigma_G^2 + \sigma_{GE}^2/E + \sigma_e^2/Er)$$

Where σ_G^2 is the genotypic variance, σ_{GE}^2 is the Genotypic by environment interaction (GEI), σ_e^2 is the error variance, E is the number of environments and r is the number of replications in each trial. The phenotypic and genotypic correlations among traits were evaluated as described by Hirel et al. (2007).

2.4 Genotypic analysis

Phenotypic data was collected on testcross hybrids whereas F₃ population lines were genotyped. Since, the single tester was used for each population, assumption is tester effect is same across lines, therefore, tester effect was not included in the model for analyses. Total DNA was extracted from bulked young leaves of the lines

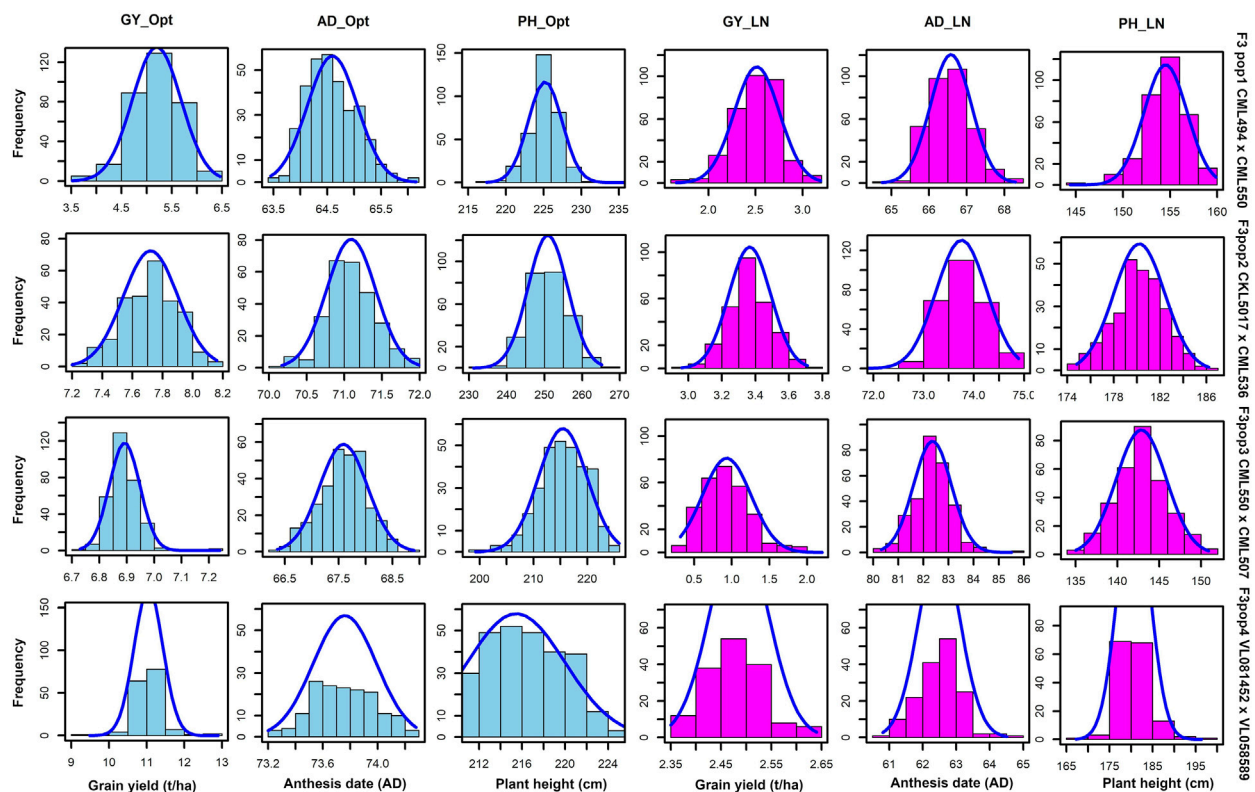


FIGURE 1

Phenotypic distribution for GY, AD and PH under optimum and low soil nitrogen conditions. The sky blue and pink colour denotes trials conducted under optimum and low nitrogen conditions, respectively. GY–Grain yield; AD - anthesis date; PH–plant height.

according to the CTAB method (CIMMYT, 2005), and the DNA quality for each sample was checked using gel-electrophoresis and spectrophotometer (NanoDrop ND8000 Thermo Scientific) before genotyping. Genotyping was performed using the Illumina MaizeSNP1500 Bead Chip evenly spaced SNP to cover the whole maize genome (Ganal et al., 2011). The above task was performed at LGC genomic labs in the United Kingdom (<https://www.lgcgroup.com/genotyping/>).

Markers which are homozygous for both the parents and polymorphic between them were retained for mapping in each population. Linkage maps in all four populations were constructed using QTL IciMapping version 4.1 software (Meng et al., 2015). Finally, we used 202, 452, 384, and 387 high-quality SNPs in F₃pop1, F₃pop2, F₃pop3, and F₃pop4, respectively. Linkage map was constructed by using these SNPs, by selecting the most significant markers using stepwise regression. A likelihood ratio test was used to calculate the logarithm of odds (LOD) for each marker at score of >3 with a 30 cM maximum distance between two loci. The recombination frequency between linked loci was transformed into cM (centi Morgan) using Kosambi's mapping function (Kosambi, 1916). BLUPs across environments were used to detect QTLs based on inclusive composite interval mapping (ICIM) for each population. Phenotypic variation explained by individual QTL and total variation explained by all QTLs together was estimated. QTL naming was done with letter "q" indicating QTL, followed by

abbreviation of trait name, the chromosome and marker position, respectively (Ribaut et al., 1997). Additive (a) and dominance (d) effects for each QTL as estimated with QTL IciMapping v.4.1 were used to calculate the ratio of dominance level ($|d/a|$). This ratio was used to classify the nature of QTL: additive (A; $0 \leq |d/a| \leq 0.2$); partially dominant (PD; $0.2 < |d/a| \leq 0.8$); dominant (D; $0.8 < |d/a| \leq 1.2$) and overdominant (OD; $|d/a| > 1.2$).

2.5 Genomic prediction

Genomic prediction (GP) analyses were conducted in R program version 4.2.1 (R Core Team 2023). GP was applied on each F₃ population to find out the prediction ability of GY and other agronomic traits evaluated in optimum and low soil N conditions. We used GP model RR-BLUP to predict the untested lines using a five-fold cross validation (Zhao et al., 2012; Crossa et al., 2017). BLUPs across environments for each of the F₃ population were used for the analysis. For GS analyses polymorphic SNPs between the parents of each population was used, like 202, 452, 384, and 387 SNPs in F₃pop1, F₃pop2, F₃pop3 and F₃pop4, respectively were used. We applied five-fold cross validations with 'within population' approach where both training and estimation set are derived from within each of biparental population. For each trait in each population, 100 iterations were done for sampling of the training and estimation sets. The prediction accuracy was

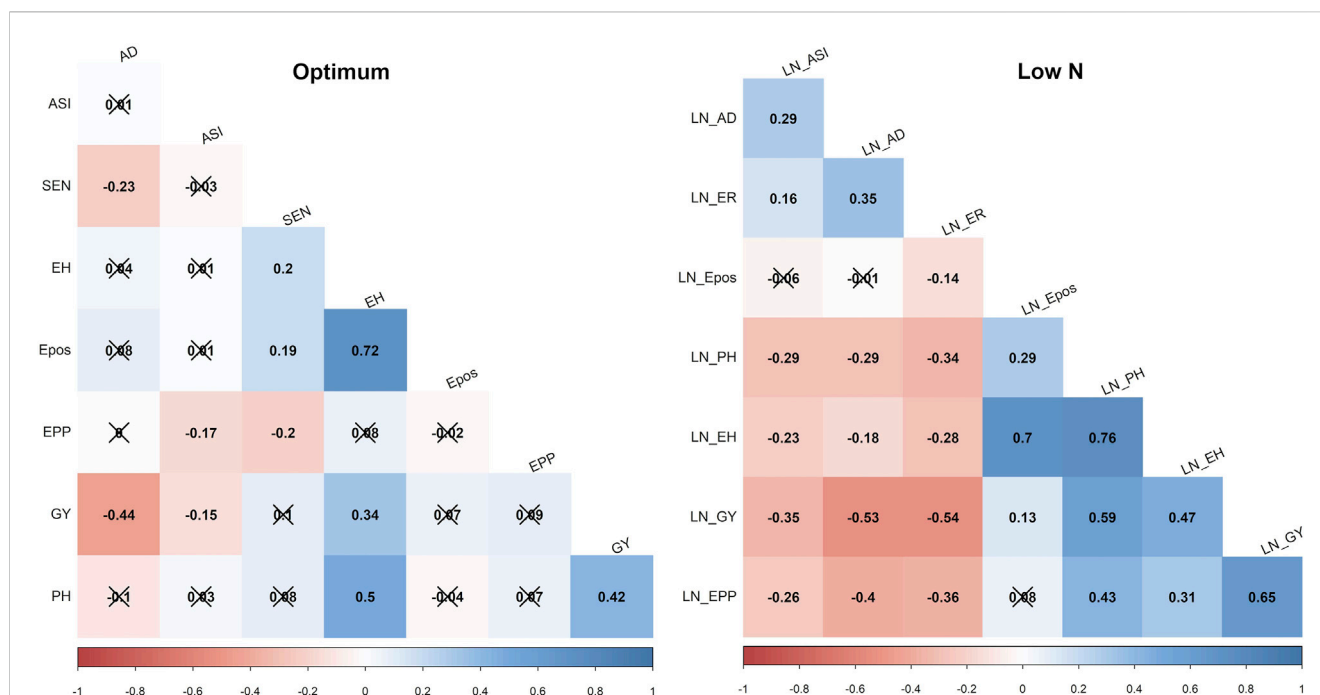


FIGURE 2

Phenotypic correlations of GY and other agronomic traits evaluated under optimum and low soil nitrogen conditions in F₃ pop1 CML494 x CML550 population. The x marks indicated values are not significant at $p < 0.05$. LN—low soil N conditions. Correlations with >0.11 , and >0.19 were significant at 0.05, and 0.01 (p) levels, respectively. GY—Grain yield; AD - anthesis date; ASI - anthesis silking interval; PH—plant height; EH, ear height; Epos—ear position; EPP—ears per plant; SEN—senescence.

calculated as the correlation between the observed and predicted breeding values divided by the square root of heritability (Dekkers, 2007).

3 Results

3.1 Phenotypic distributions and correlation of traits

GY, ASI, AD, PH, EPO, and EH varied widely across the two nitrogen regimes (i.e., optimum and low nitrogen environments). The extent of variation, however, differed between the four biparental maize populations. Low nitrogen conditions increased the trial mean for AD and decreased the trial mean for GY and PH across the four F₃ populations tested in this study (Figure 1, Supplementary Table S1). GY showed a negative genotypic correlation with AD and ASI across the two nitrogen regimes (Figure 2). Moreover, consistent positive genetic correlations for GY were recorded with EH, EPO, and PH.

3.2 Trial mean, genetic variance, and heritability of traits

Population 1 (CML494x CML550): Across the study sites, average GY was 5.22 t/ha and 2.52 t/ha under optimum and low soil nitrogen conditions, respectively (Table 2). A margin of 2-day extension in AD was observed under low soil nitrogen conditions.

The genotypic and genotype \times environment interaction effects were significant ($p \leq 0.05$) for GY. Under optimum conditions, moderate heritabilities across environments were recorded for AD (0.55), GY (0.54), and EH (0.51). Similarly, AD (0.62), GY (0.58), and EH (0.55) exhibited slightly higher heritability under low nitrogen conditions. The lowest heritabilities under optimum nitrogen conditions were exhibited by PH (0.36) and EPO (0.42) traits. ASI (0.18), PH (0.39), and EPO (0.48) recorded the lowest heritability values under low nitrogen conditions.

Population 2 (CKL5017x CML536): Under optimum conditions (Table 2), moderate heritability was observed for PH (0.68) and EH (0.49) whilst GY (0.28) recorded low heritability across the studied sites. Under low nitrogen conditions, all the studied traits recorded low heritabilities with GY pegged at 0.31. The genotypic and genotype \times environment interaction effects were significant ($p \leq 0.05$) for GY. Across the studied regions, the mean GY was 7.72 t/ha and 3.37 t/ha under optimum and low nitrogen conditions, respectively.

Population 3 (CML550x CML507): Low heritabilities were recorded across the tested traits under both optimum and low nitrogen conditions (Table 2). GY heritabilities were 0.31 and 0.29 under optimum and low nitrogen conditions, respectively. The genotypic and genotype \times environment interaction effects were significant ($p \leq 0.05$) for GY. In this population, the mean GY was 6.89 t/ha and 0.94 t/ha under optimum and low nitrogen conditions, respectively.

Population 4 (VL081452x VL058589): All traits tested using population 4 under optimum conditions recorded low heritabilities (Table 2) including GY (0.27). Under low nitrogen conditions, AD (0.58) exhibited moderate heritability whilst GY

TABLE 2 Trait means, heritability and variance components of GY and other traits evaluated under optimum and low nitrogen conditions in four F₃ populations across multiple environments.

	Trait	Mean	σ^2_G	$\sigma^2_{G \times E}$	σ^2_e	H ²
F3 pop1 CML494×CML550						
Optimum	GY	5.22	0.44**	0.51**	1.25	0.54
	AD	64.58	0.42**	0.26**	1.53	0.55
	PH	225.27	17.10**	12.19**	161.99	0.36
	EH	107.94	15.09**	1.23**	82.84	0.51
	EPO	0.46	0.0002*	0.0000	0.0012	0.42
Low N	GY	2.52	0.11**	0.10**	0.41	0.58
	AD	66.57	0.53**	0.30**	1.97	0.62
	ASI	−0.51	0.08*	0.06*	1.35	0.18
	PH	154.64	14.92**	23.10**	137.33	0.39
	EH	67.05	11.84**	7.55**	62.73	0.55
	EPO	0.41	0.0002*	0.00	0.002	0.48
F3 pop 2 CKL5017×CML536						
Optimum	GY	7.72	0.12**	0.15**	1.60	0.30
	AD	71.10	0.42**	0.03**	2.34	0.41
	PH	250.79	45.01**	4.99**	74.41	0.68
	EH	83.98	33.01**	0.00	129.88	0.68
	EPO	0.46	0.0003**	0.00	0.003	0.41
Low N	GY	3.37	0.06**	0.01*	0.84	0.31
	AD	73.75	0.70**	0.29**	3.49	0.41
	PH	180.17	16.61**	5.91**	185.53	0.34
	EH	83.98	20.01**	0.00	126.88	0.49
	EPO	0.46	0.0003*	0.00	0.004	0.41
F3 pop 3 CML550×CML507						
Optimum	GY	6.89	0.11*	0.14**	1.24	0.31
	AD	67.59	0.50**	0.16**	2.32	0.43
	PH	215.46	43.72**	4.35**	156.63	0.51
	EH	106.11	14.93**	4.33**	94.40	0.37
Low N	GY	0.94	0.03*	-	0.13	0.29
	AD	82.37	1.10**	-	1.69	0.57
	PH	142.88	26.73**	-	83.97	0.39
	EH	63.55	10.25**	-	34.65	0.37
F3 pop 4 VL081452×VL058589						
Optimum	GY	11.05	0.53**	-	2.82	0.27
	AD	73.76	0.32*	-	2.86	0.18
	PH	262.81	12.96**	-	194.11	0.14
	EH	2.01	1.08**	-	7.08	0.20

(Continued on following page)

TABLE 2 (Continued) Trait means, heritability and variance components of GY and other traits evaluated under optimum and low nitrogen conditions in four F₃ populations across multiple environments.

	Trait	Mean	σ^2_G	$\sigma^2_{G \times E}$	σ^2_e	H ²
Low N	GY	2.49	0.08*	0.05*	0.42	0.39
	AD	62.49	0.70**	0.27**	1.47	0.58
	ASI	1.27	0.26**	0.08**	3.17	0.21
	PH	173.09	40.44**	0.00	93.40	0.63

*, ** Significant at $p < 0.05$ and $p < 0.01$ levels, respectively. Low N—low soil N conditions. GY, Grain yield; AD, anthesis date; ASI, anthesis silking interval; PH, plant height; EH, ear height; EPO, ear position; σ^2_G , genotypic variance; $\sigma^2_{G \times E}$, genotypic \times environment interaction variance; σ^2_e , residual error variance; H² - heritability.

(0.39) and ASI (0.21) had the lowest heritabilities. The genotypic and genotype \times environment interaction effects were significant ($p \leq 0.05$) for GY. Across the study sites, average yields were 11.05 t/ha and 2.49 t/ha under optimum and low nitrogen conditions, respectively.

3.3 QTLs mapping in four F₃ maize populations

The study obtained a total of 906.83, 1,650.28, 1857.12, and 2169.97 cM from 202, 452, 384, and 387 polymorphic SNPs for F₃ population 1, 2, 3, and 4, respectively (Table 3). The mean distance between adjacent markers was lowest in population 2 (3.65 cM) and highest for population 4 (5.61 cM). Across the ten maize chromosomes, a total of 91 significant QTLs were identified for GY (11), AD (26), PH (19), ASI (4), EH (24), and EPO (7) under optimum (52) and low nitrogen (39) conditions (Tables 4–7). The identified QTLs were distributed across 10 chromosomes. Across environments, chromosomes 1 (16) and 8 (14) had the highest number of QTLs, whilst chromosomes 7 (5) and 5 (6) had the lowest number of QTLs. The identified QTLs were distributed as 19, 36, 18, and 36 for populations 1, 2, 3, and 4, respectively. Proportional phenotypic variance for each QTL ranged from 0.7% (AD for population 2 under optimum nitrogen) to 15.22% (AD for population 4 under low nitrogen).

Eleven QTLs were identified for GY under optimum (7) and low soil nitrogen (4) conditions distributed across all chromosomes except chromosome 4, 5, 6, and 8 (Tables 4–8). Across all genotypes and nitrogen regimes, no common QTL was identified. QTLs associated with GY under low nitrogen were detected in chromosomes 1 (population 2, 3, and 4) and 10 (population 3). On the other hand, QTL underlying GY under optimum nitrogen conditions were observed on chromosomes 1 (population 1), 2 (population 4), 8 (population 2), 9 (populations 1 and 3), and 10 (population 4). Under optimum conditions, the total phenotypic variance explained (TPVE) by all QTL was 11.58%, 16.52%, 12.15%, and 7.68% for populations 1, 2, 3, and 4, respectively. TPVE by all QTL under low nitrogen conditions was 6.20% (population 2), 5.12% (population 3) and 19.16% (population 4). Among the QTL detected for GY, we observed additive, dominance, partial dominance and over dominance effect QTL across 4 populations.

For AD, 26 significant QTLs were identified across the two nitrogen regimes. The QTLs were detected on chromosomes 1, 2, 3, 4, 7, and 10 under low nitrogen and on chromosomes 1, 2, 3, 4, 5, 6, 8, and 9 under optimum nitrogen conditions. No common QTL for the two nitrogen regimes was identified across the studied F₃ populations. TPVE by all QTL under optimum conditions was

8.09%, 18.89%, 11.80% and 12.64% for populations 1, 2, 3 and 4, respectively. Under low nitrogen conditions, TPVE by all QTL was 7.70% (population 1), 18.55% (population 2), 7.45% (population 3) and 35.24% (population 4). The nature of the QTLs classified a few as additive, dominant, partial dominant and overdominance groups.

Four QTLs were identified for ASI under low nitrogen conditions. These were detected on chromosomes 1, 3, 4, and 6. As in the case of GY and AD, no common QTL was identified. TPVE by all QTLs under low nitrogen conditions was 6.91% (population 1) and 16.73% (population 4).

For EH, 24 significant QTLs were detected with 9 of those under low soil nitrogen conditions. The QTLs underlying EH under low soil nitrogen conditions were detected in chromosomes 1, 2, 3, 5, 6, 7, 8, and 9. Under optimum conditions, the QTLs were found in all 10 maize chromosomes except for chromosome 7. Common QTL for the two nitrogen regimes was not identified. TPVE by all QTLs under optimum conditions was 23.20%, 23.35%, 22.25% and 15.26% for populations 1, 2, 3, and 4, respectively. Under low soil nitrogen conditions, TPVE by all QTL was 4.87% (population 1), 14.42% (population 2), and 13.43% (population 3).

Significant QTLs underlying EPO were recorded as 7 under both nitrogen regimes. One QTL (Chromosome 6: *qEPO6_157*) was found in both low and optimum nitrogen conditions. Under low soil nitrogen management, QTLs were detected in chromosomes 3, 4, and 6. On the other hand, chromosomes 1, 3, 6, and 8 housed QTL underlying EPO under optimum conditions. Under optimum conditions, the TPVE by all QTL was 15.56%, and 11.13% for populations 1 and 2, respectively. TPVE by all QTLs under low nitrogen conditions was 11.56% (population 1) and 17.98% (population 2).

For PH, 19 significant QTLs were identified across the studied genotypes and nitrogen regimes. Nine of those were identified under low nitrogen conditions in chromosomes 1, 2, 3, 6, 7, and 9. Under optimum conditions, significant QTLs were detected in chromosomes 1, 4, 6, 8, and 9. No common QTL for the two nitrogen regimes was identified across the studied F₃ populations. TPVE by all QTLs under optimum conditions was 6.00%, 25.07%, and 25.81% for populations 1, 2, and 3, respectively. Under low nitrogen conditions, TPVE by all QTL was 9.11% (population 1), 18.76% (population 2), 7.27% (population 3) and 6.34% (population 4).

3.4 Overlapping QTL for each trait evaluated under low and optimum nitrogen conditions

The identification of common QTLs is crucial in targeting markers that can be used in breeding for improved nitrogen stress

TABLE 3 Linkage map information for the four F3 populations used for QTL analysis.

Chromosome	F3 pop1 CML494×CML550		F3 pop 2 CKL5017×CML536		F3 pop 3 CML550×CML507		F3 pop 4 VL081452×VL058589	
	SNPs	Distance (cM)	SNPs	Distance (cM)	SNPs	Distance (cM)	SNPs	Distance (cM)
1	28	114.37	73	304	63	279.88	74	392.13
2	15	101.71	57	234	42	234.4	39	207.9
3	33	122.62	45	180.5	43	232.17	33	235.77
4	26	83.39	55	195.97	35	220.4	55	235.5
5	25	125.19	53	137.22	54	213.61	50	225.05
6	17	101.46	32	121.78	33	148.3	33	243.07
7	20	47	30	91.55	23	71.67	27	86.99
8	18	55.6	46	172.03	36	108.61	30	174.8
9	13	80.04	27	115.8	23	151.78	20	230.13
10	7	75.45	34	97.43	32	196.3	26	138.33
Total	202	906.83	452	1,650.28	384	1857.12	387	2169.67
Avg dist		4.49		3.65		4.84		5.61

tolerance and NUE. This study identified several common QTLs across the studied F3 populations and nitrogen regimes (Tables 4–8). Under optimum conditions, one QTL (*qAD03_158*) for AD was overlapping with one QTL under low soil nitrogen conditions on chromosome 3 (157.97–159.08 Mbp). For ASI and PH, no common QTL was identified across the two nitrogen regimes. Under low soil nitrogen conditions, one QTL for EH, *qEH01_21* (16.42–42.32 Mbp), was overlapping with two QTLs under optimum conditions (from 20.99 to 21.87 Mbp and 16.42 to 42.32 Mbp) on chromosome 1. Similarly, under low nitrogen, the second QTL for EH (*qEH09_108*: 107.26 to 113.71 Mbp) overlapped with one QTL under optimum conditions (28.74–119.42 Mbp) on chromosome 9. One common QTL for EPO, *qEP06_157*, was found under both low and optimum nitrogen conditions (156.17–159.55 Mbp and 156.17 to 159.55 Mbp) on chromosome 6. Under optimum conditions, one QTL for GY (*qGY10_130*) was found on chromosome 10 in both population 3 (125.47–131.38 Mbp) and population 4 (46.11–130.61 Mbp). A similar observation was made for one QTL underlying EH (*qEH09_108*) which was seen on chromosome 9 under optimum nitrogen conditions in population 3 (28.74–119.42 Mbp) and low nitrogen conditions in population 2 (107.26–113.71 Mbp). This study found no QTL correspondence among the studied traits across the nitrogen regimes.

3.5 Genomic prediction correlations for traits evaluated under low and optimum nitrogen conditions

To assess the potential breeding value of GY and other agronomic traits under optimum and low soil N management, GS analysis was performed on all four populations using NUE-associated traits and genotypic markers with RR-BLUP. The RR-BLUP prediction correlations were 0.35, 0.41, 0.06, and 0.35 in populations 1, 2, 3, and 4, respectively (Figure 3; Table 9) under optimum, whereas under low soil N management, the correlations were reduced to 0.21, 0.41, 0.04, and –0.02 in population 1, 2, 3, and 4, respectively. For AD, prediction correlations were 0.31, 0.63, 0.12, and 0.37 under optimum and 0.26, 0.43, 0.29, and 0.36 under low soil N management, respectively (Figure 3; Table 9). For PH, prediction correlations were 0.35, 0.52, 0.34, and 0.18 under optimum, and 0.44, 0.40, 0.15, and 0.10 under low soil N management in populations 1, 2, 3, and 4, respectively. Overall, the prediction correlations were higher in population 2 and lower in population 4 for all the traits under both optimum and low soil N management.

4 Discussion

4.1 Impact of low nitrogen stress on GY and other associated traits

The role of nitrogen stress in yield reduction and overall crop development cannot be overstated. An in-depth understanding of GY and other related agronomic traits is critical for the evidence-based development of varieties that are tolerant to low soil nitrogen stress (Derera et al., 2008). However, phenotypic characterization of such complex traits is very challenging due to unpredictable

TABLE 4 Genetic characteristics of QTLs detected for GY and associated traits under optimum and low-nitrogen stress in F₃ population 1.

Mgmt	Trait	Chr	QTL name	Position (cM)	Left marker	Right marker	Physical position (Mbp)	LOD	PVE (%)	TPVE (%)	Add	Dom	QTL
F3 pop1 CML494×CML550													
Opt	GY	1	<i>qGY01_21</i>	83	PZA02487.1	PZA00425.11	20.99–21.87	4.94	6.03	11.58	−0.18	0.009	A
		9	<i>qGY09_30</i>	60	PZB01110.6	PHM13183.12	29.66–76.42	5.39	6.91		0.19	0.029	A
	AD	1	<i>qAD01_246</i>	27	PZB00895.1	PZA01588.1	245.60–257.19	3.45	5.17	8.09	−0.14	−0.063	PD
	PH	7	<i>qPH07_88</i>	13	PZB00174.1	PZA01690.7	87.78–105.88	3.32	6.00	6.00	−2.20	−1.069	PD
	EH	1	<i>qEH01_21</i>	83	PZA02487.1	PZA00425.11	20.99–21.87	3.42	5.07	23.20	−0.71	−0.071	A
		4	<i>qEH04_115</i>	46	PZA00704.1	PZA03564.1	114.90–119.66	3.53	5.30		−0.72	−0.039	A
		8	<i>qEH08_60</i>	23	PZA02683.1	PZA00379.2	59.15–74.72	5.42	8.46		−0.87	0.216	PD
		10	<i>qEH10_95</i>	75	PHM229.15	PZB01358.1	92.58–107.42	3.40	5.22		0.74	−0.037	A
	EPO	6	<i>qEP06_157</i>	0	PHM5361.13	PZA00889.2	156.17–159.55	4.68	6.16	15.56	0.00	0.002	PD
		8	<i>qEP08_60</i>	23	PZA02683.1	PZA00379.2	59.15–74.72	9.39	12.76		0.00	0.003	PD
Low N	AD	8	<i>qAD08_128</i>	15	PZB01454.1	PHM15744.10	127.95–138.96	3.05	5.00	7.70	−0.16	0.023	A
	ASI	1	<i>qASI01_21</i>	96	PZA00425.11	PZA00566.5	9.53–20.99	3.81	3.96	6.91	0.03	−0.045	OD
		3	<i>qASI03_171</i>	88	PHM17210.5	PZA00538.15	170.91–199.68	3.02	3.07		0.03	0.027	D
		6	<i>qASI06_71</i>	83	PZA02606.1	PZA01960.1	70.29–156.68	3.03	6.80		0.02	0.075	OD
	PH	1	<i>qPH01_22</i>	73	PZA01963.15	PZA02487.1	21.86–191.32	4.46	9.01	9.11	−1.06	0.692	PD
		6	<i>qPH06_120</i>	40	PZA02673.1	PZA00473.5	117.06–135.75	3.50	3.19		−0.53	0.633	D
		7	<i>qPH07_120</i>	22	PZA01542.1	PZA02854.13	112.40–120.22	4.13	3.76		−0.74	−0.146	A
	EH	7	<i>qEH07_106</i>	20	PZA01690.7	PZA01542.1	105.88–112.40	3.50	4.87	4.87	−0.72	−0.222	PD
	EPO	4	<i>qEP04_115</i>	46	PZA00704.1	PZA03564.1	114.90–119.66	3.08	3.80	11.56	−0.01	−0.003	PD
		6	<i>qEP06_157</i>	1	PHM5361.13	PZA00889.2	156.17–159.55	5.82	7.76		0.01	0.006	D

Opt, Optimum; Low N- low soil N stress management; GY, Grain yield; AD, anthesis date; ASI, anthesis silking interval; PH, plant height; EH, ear height; EPO, ear position. The italic values refer to the names of identified QTLs.

environmental and edaphic conditions (Liu et al., 2010). Importantly, selection through morphological characterization under nitrogen stress conditions is very inconsistent and will not give correct phenotypic data thereby derailing progress in maize breeding schemes. In contrast, the integration of precise genomic tools coupled with conventional breeding can accelerate the development of nitrogen-stress adaptive cultivars with high yields. In this research, two nitrogen regimes (i.e., low, and optimum nitrogen) served as the basis for the identification of QTLs related to GY and other agronomic traits (i.e., AD, ASI, PH, EH, and EPO) and their cross-cutting heritabilities in four F₃ maize populations.

The target traits measured followed a normal distribution indicating that the studied traits are quantitative in nature and the tested F₃ populations are stable and suitable for QTL mapping. The mean GY reduction in this study was 69.82% under low nitrogen stress. This is very close to the 71% yield reduction recorded in the study by Ertiro T. B. et al. (2020). Yield reduction under low soil nitrogen stress demonstrates the importance of nitrogen in the growth and development of maize (Ertiro T. B. et al., 2020). Numerous studies have linked a reduction

in yield under low-N conditions to decreased kernel number due to abortion (Bänziger et al., 1997; Agrama et al., 1999; Ribaut et al., 2007). Low soil nitrogen environments decrease the number of kernels and ears per plant, decrease the chlorophyll concentration in the ear leaf, and lowers PH by roughly half (Ribaut et al., 2007).

Genotypic and genotype × environment interaction effects were significant for GY across the tested nitrogen regimes. Generally, across the studied environments, the most responsive trait was AD. Our results showed that low nitrogen stress extends the AD in maize. Furthermore, low nitrogen stress increased genetic variance for AD across the studied genotypes. This can be viewed as an indication of the adaptive nature of this secondary trait.

Across the four F₃ populations, there was a consistent moderate to low heritability on all the studied traits across the two nitrogen regimes. For all traits, heritability was shown to decline under low nitrogen stress. The heritabilities of almost all the studied traits were moderate to low. Similar trends were also reported in earlier studies with DH populations (Ertiro et al., 2020b; Ertiro et al., 2022) and association panel (Ertiro et al., 2020a) which were evaluated under optimum and low soil N conditions. Despite this, the heritabilities were significant enough to facilitate indirect selection for increased

TABLE 5 Genetic characteristics of detected QTLs for GY and associated traits under optimum and low-nitrogen stress in F3 population 2.

Mgmt	Trait	Chr	QTL name	Position (cM)	Left marker	Right marker	Physical position (Mbp)	LOD	PVE (%)	TPVE (%)	Add	Dom	QTL
F3 pop 2 CKL5017×CML536													
Opt	GY	3	<i>qGY03_10</i>	144	PZD00038.2	PHM4621.57	4.65–159.08	3.41	4.08	16.52	−0.15	−0.047	PD
		8	<i>qGY08_127</i>	100	PZA00460.8	PHM4203.11	126.42–156.31	4.84	4.15		0.15	0.059	PD
	AD	1	<i>qAD01_215</i>	188	PHM759.24	PZA00664.3	213.95–216.05	4.96	0.70	18.89	0.40	0.002	A
		2	<i>qAD02_55</i>	95	PHM3055.9	PHM4259.5	51.96–186.31	9.84	2.20		−4.19	−4.208	D
		2	<i>qAD02_145</i>	151	PHM5060.12	PZA03211.6	142.76–184.40	17.23	2.23		−0.06	8.428	OD
		3	<i>qAD03_158</i>	180	PZA00186.4	PZA01396.1	157.97–159.08	14.47	2.23		4.20	−3.541	D
		4	<i>qAD04_14</i>	50	PHM16788.6	PHM 2006.57	13.89–171.71	9.06	2.20		4.14	−4.264	D
		5	<i>qAD05_10</i>	18	PZA00963.3	PZA00818.1	1.1–208.34	14.45	2.23		−4.19	−4.629	D
		6	<i>qAD06_72</i>	1	PHM2898.24	PZD00072.2	71.99–72.48	14.32	2.21		4.20	−3.821	D
		9	<i>qAD09_23</i>	114	PHM4720.12	PZA03671.1	22.54–101.88	12.89	2.20		0.04	8.352	OD
		10	<i>qAD10_130</i>	8	PHM3736.11	PZA03605.1	125.47–131.38	15.65	2.29		−0.83	9.220	OD
	PH	1	<i>qPH01_45</i>	149	PZA00962.1	PZA00939.1	42.32–104.96	4.84	9.75	25.07	2.15	0.631	PD
		6	<i>qPH06_155</i>	106	PHM4503.25	PZA02815.25	154.00–160.83	3.05	4.26		1.48	0.142	A
		8	<i>qPH08_104</i>	32	PZA01972.14	PZA02566.1	103.35–104.22	3.38	8.35		2.11	0.302	A
		8	<i>qPH08_130</i>	67	PZA00460.8	PHM4203.11	126.42–156.31	5.44	7.74		2.05	0.007	A
	EH	1	<i>qEH01_21</i>	136	PZA02393.2	PZA00962.1	16.42–42.32	4.88	8.55	26.35	1.68	0.895	PD
		1	<i>qEH01_220</i>	189	PZA00664.3	PHM4992.10	216.05–225.87	3.18	6.72		1.59	−0.229	A
		3	<i>qEH03_50</i>	48	PZA00581.3	PZA02645.2	46.96–89.25	4.75	6.45		1.63	0.490	PD
		8	<i>qEH08_130</i>	67	PZA00460.8	PHM4203.11	126.42–156.31	3.20	4.03		1.20	0.521	PD
	EPO	1	<i>qEP01_214</i>	188	PHM759.24	PZA00664.3	213.95–216.05	3.09	3.22	11.13	0.00	0.000	A
		3	<i>qEP03_153</i>	31	PZA00667.2	PHM9914.11	152.70–154.23	3.77	7.70		0.00	0.000	A
Low N	GY	1	<i>qGY01_60</i>	230	PHM4752.14	PHM574.14	59.27–288.94	3.43	6.27	6.20	0.04	0.026	PD
	AD	2	<i>qAD02_52</i>	94	PHM3055.9	PHM4259.5	51.96–186.31	4.64	1.52	18.55	1.16	1.128	D
		2	<i>qAD02_53</i>	123	PHM4259.5	PZA01991.3	51.96–221.13	4.68	1.52		1.15	1.152	D
		3	<i>qAD03_120</i>	41	PZA00363.7	PZA00707.9	98.45–120.53	5.70	1.30		0.27	0.002	A
		3	<i>qAD03_158</i>	180	PZA00186.4	PZA01396.1	157.97–159.08	4.49	1.15		−1.44	1.717	D
		4	<i>qAD04_172</i>	52	PHM 2006.57	PZA01905.12	171.17–251.33	6.29	1.52		−0.74	0.366	PD
		4	<i>qAD04_207</i>	194	PHM3587.6	PHM14055.6	206.77–208.12	6.92	1.45		0.01	−2.298	OD
		7	<i>qAD07_82</i>	90	PZA01933.3	PHM3435.6	80.61–141.80	6.43	1.50		0.05	−2.262	OD
		8	<i>qAD08_126</i>	170	PHM4203.11	PZA00766.1	126.42–126.71	4.46	1.26		−1.44	1.757	OD
		10	<i>qAD10_130</i>	8	PHM3736.11	PZA03605.1	125.47–131.13	7.73	1.24		0.07	−2.885	OD
	PH	1	<i>qPH01_21</i>	137	PZA02393.2	PZA00962.1	16.42–42.32	4.45	6.45	18.76	0.84	−0.081	A
		3	<i>qPH03_175</i>	18	PZA00892.5	PZA03735.1	173.26–192.67	4.13	5.32		−0.76	0.494	PD
		9	<i>qPH09_88</i>	22	PHM1766.1	PZA03235.1	86.54–107.26	3.21	4.32		0.71	−0.517	PD
		9	<i>qPH09_86</i>	32	PZA03235.1	PZA00225.8	76.31–86.54	4.02	4.58		−0.76	−0.050	A

(Continued on following page)

TABLE 5 (Continued) Genetic characteristics of detected QTLs for GY and associated traits under optimum and low-nitrogen stress in F3 population 2.

Mgmt	Trait	Chr	QTL name	Position (cM)	Left marker	Right marker	Physical position (Mbp)	LOD	PVE (%)	TPVE (%)	Add	Dom	QTL
	EH	1	<i>qEH01_21</i>	140	PZA02393.2	PZA00962.1	16.42–42.32	3.04	3.34	14.42	0.81	–0.257	PD
		3	<i>qEH03_155</i>	26	PHM17210.5	PZA00667.2	154.23–170.91	7.63	10.23		1.41	0.700	PD
		3	<i>qEH03_155</i>	33	PZA00667.2	PHM9914.11	152.70–170.91	5.32	6.86		–1.25	0.350	PD
		9	<i>qEH09_108</i>	21	PZA00323.3	PHM1766.1	107.26–113.71	3.14	3.39		0.84	–0.713	D
	EPO	3	<i>qEP03_156</i>	26	PHM17210.5	PZA00667.2	154.23–170.91	9.01	8.12	17.98	0.01	0.002	PD
		3	<i>qEP03_155</i>	33	PZA00667.2	PHM9914.11	152.70–170.91	4.70	4.28		0.00	0.001	PD

Opt, Optimum; Low N- low soil N stress management; GY, Grain yield; AD, anthesis date; ASI, anthesis silking interval; PH, plant height; EH, ear height; EPO, ear position. The italic values refer to the names of identified QTLs.

TABLE 6 Genetic characteristics of detected QTLs for GY and associated traits under optimum and low-nitrogen stress in F3 population 3.

Mgmt	Trait	Chr	QTL name	Position (cM)	Left marker	Right marker	Physical position (Mbp)	LOD	PVE (%)	TPVE (%)	Add	Dom	QTL
F3 pop 3 CML550×CML507													
Opt	GY	9	<i>qGY09_25</i>	126	PZB01110.6	sh1.11	17.01–29.66	3.20	3.58	12.15	0.02	–0.003	A
		10	<i>qGY10_130</i>	7	PHM3736.11	PZA03603.1	125.47–131.38	3.87	6.21		0.00	0.035	OD
	AD	3	<i>qAD03_175</i>	2	PZA01962.12	PZA00892.5	170.91–192.67	3.71	6.09	11.8	0.15	0.003	A
		5	<i>qAD05_72</i>	61	PHM13675.17	PZA00261.6	71.08–79.98	3.40	5.07		–0.11	0.115	D
		8	<i>qAD08_70</i>	19	PZA01363.2	PZA03012.7	65.14–107.76	4.49	7.10		–0.17	–0.065	PD
	PH	1	<i>qPH01_210</i>	153	PZA00664.3	PHM4997.11	6.02–216.05	4.71	5.59	25.81	–1.58	0.298	A
		1	<i>qPH01_05</i>	181	PHM4997.11	PZA02129.1	3.83–6.02	3.49	3.87		1.20	0.271	PD
		4	<i>qPH04_40</i>	166	PZA03597.1	PZA00541.1	36.33–61.17	3.97	4.32		1.22	–1.050	D
		8	<i>qPH08_125</i>	43	PZA01049.1	PZA00770.1	122.03–127.01	4.64	5.31		1.58	0.312	PD
		9	<i>qPH09_25</i>	126	PZB01110.6	sh1.11	17.01–29.66	4.70	5.16		1.56	–0.347	PD
	EH	5	<i>qEH05_165</i>	95	PZA00148.3	ae1.7	163.41–167.39	3.49	4.67	22.25	–0.71	–0.126	A
		8	<i>qEH08_125</i>	43	PZA01049.1	PZA00770.1	122.03–127.01	5.37	7.48		0.82	0.330	PD
		9	<i>qEH09_108</i>	123	PHM 1911.173	PZB00544.2	28.74–119.42	5.19	7.62		0.79	–0.520	PD
Low N	GY	1	<i>qGY01_200</i>	22	PHM 1968.22	PZA01294.2	63.80–202.86	3.07	5.21	5.12	0.02	–0.014	PD
		10	<i>qGY10_120</i>	34	PZA01642.1	PHM15868.56	12.42–120.37	3.45	1.16		–0.33	–0.333	D
	AD	3	<i>qAD03_175</i>	0	PZA01962.12	PZA00892.5	170.91–192.67	3.44	5.74	7.45	0.20	–0.174	D
	PH	2	<i>qPH02_25</i>	82	PHM4425.25	PZA02378.7	19.52–33.65	3.02	5.00	7.27	–2.95	0.498	A
	EH	2	<i>qPH02_80</i>	109	PHM10321.11	PZA01280.2	61.36–143.35	4.07	3.87	13.43	0.57	–0.199	PD
		5	<i>qPH05_10</i>	4	PZA01327.1	PZA02367.1	8.62–15.22	3.03	5.14		0.65	1.150	OD
		6	<i>qPH06_25</i>	86	PZA02815.25	PHM15961.13	8.62–160.83	4.54	7.18		–0.80	–0.006	A
		8	<i>qPH08_140</i>	44	PZB01454.1	PZA00838.2	138.96–151.85	4.52	5.04		–0.71	0.316	PD

Opt, Optimum; Low N- low soil N stress management; GY, Grain yield; AD, anthesis date; PH, plant height; EH, ear height. The italic values refer to the names of identified QTLs.

TABLE 7 Genetic characteristics of detected QTLs for GY and associated traits under optimum and low-nitrogen stress in F₃ population 4.

Mgmt	Trait	Chr	QTL name	Position (cM)	Left marker	Right marker	Physical position (Mbp)	LOD	PVE (%)	TPVE (%)	Add	Dom	QTL
F3 pop 4 VL081452×VL058589													
Opt	GY	2	<i>qGY02_25</i>	186	PZA01820.1	PZA02264.5	2.55–46.62	3.26	2.14	7.68	0.05	1.243	OD
		10	<i>qGY10_130</i>	137	PZA00062.4	PZA00409.17	46.11–130.61	3.49	3.17		−0.01	1.381	OD
	AD	8	<i>qAD08_25</i>	139	PZA01196.2	PZA03612.1	24.45–121.52	3.79	10.73	12.64	−0.04	0.422	OD
	EH	2	<i>qER02_03</i>	173	PZA00680.3	PHM5535.8	1.00–3.46	6.24	3.05	15.26	0.07	0.763	OD
		2	<i>qER02_185</i>	196	PZA02170.1	PHM5060.12	184.40–231.71	6.06	2.87		0.00	0.823	OD
		5	<i>qER05_55</i>	194	PZA02207.1	PHM2769.43	51.25–59.62	5.78	3.77		0.01	0.624	OD
		5	<i>qER05_190</i>	213	PHM2769.43	PHM7908.25	59.62–191.25	10.17	3.90		0.02	0.782	OD
		6	<i>qER06_140</i>	51	PZB01569.7	PZA02478.7	134.48–153.63	3.20	2.72		0.38	−0.369	D
		8	<i>qER08_30</i>	160	PZA01196.2	PZA03612.1	24.45–121.52	5.89	3.39		−0.03	0.728	OD
		10	<i>qER10_80</i>	10	PHM3309.8	PZA01451.1	7.12–128.41	3.43	3.00		0.30	−0.397	OD
		10	<i>qER10_130</i>	135	PZA00062.4	PZA00409.17	46.11–130.61	3.88	1.95		−0.05	0.846	OD
Low N	GY	1	<i>qGY01_125</i>	204	PZA02135.2	PZA01254.2	105.20–155.34	3.65	6.42	19.16	0.02	0.005	PD
	AD	1	<i>qAD01_200</i>	310	PZA01216.1	PZA03265.3	190.67–203.29	3.91	4.99	35.24	−0.41	−0.109	PD
		2	<i>qAD02_30</i>	188	PZA01820.1	PZA02264.5	2.55–46.62	3.65	4.77		0.15	1.752	OD
		7	<i>qAD07_125</i>	41	PHM9162.135	PZA02959.14	120.22–133.75	11.05	15.22		0.71	0.077	A
		10	<i>qAD10_125</i>	25	PHM3309.8	PZA01451.1	7.12–128.41	3.20	10.51		0.57	0.344	PD
	PH	2	<i>qPH02_230</i>	5	PZD00022.5	PZA02727.1	228.51–233.61	3.84	7.10	6.34	1.33	−2.460	OD
	ASI	4	<i>qASI04_05</i>	170	PHM3963.33	PHM3301.28	5.54–5.81	4.00	13.58	16.73	−0.03	−0.103	OD

Opt, Optimum; Low N- low soil N stress management; GY, Grain yield; AD, anthesis date; ASI, anthesis silking interval; PH, plant height; EH, ear height. The italic values refer to the names of identified QTLs.

GY under low nitrogen stress. By comparison, PH had low heritabilities under both low and optimum nitrogen conditions.

The pairwise correlation showed a strong negative correlation for GY with AD and ASI across the studied nitrogen regimes. However, other agronomic traits showed positive correlations with GY. These include PH, EH and EPO under both low and optimum soil nitrogen conditions. Similar correlation trends were also reported by earlier studies with maize hybrid and inbred line trials under low soil nitrogen conditions (Ertiro T. B. et al., 2020; Ertiro et al., 2022). PH, EH and EPO had a positive correlation with GY suggesting that improvement of these traits lead to improved maize varieties with high-yielding potentials. Furthermore, these correlations suggest the high odds of tightly linked loci controlling low-N tolerance through the coordinated expression of loci controlling these traits.

4.2 QTL mapping under low and optimum nitrogen conditions

NUE traits are highly complex, the advent of high-density marker data has made it feasible to dissect their underlying genetics. The detection of QTLs underlying GY and other

associated traits under different nitrogen regimes is integral for scaling breeding initiatives targeting the development of low nitrogen tolerant maize varieties. A total of 91 significant QTLs were identified in this study for six traits under low and optimum nitrogen management (Tables 5, 6, 7, and 8). The disparity in the number of QTLs under optimum (52) and low (39) indicates the effect of genetic variance across the environments. Some of these were found across management conditions and populations. Across the two nitrogen regimes, chromosomes 1 (16) and 8 (14) had the highest number of QTLs whilst chromosomes 7 (5) and 5 (6) had the lowest. Chromosomes with a high number of QTLs can be targeted for future research studies focusing on improving GY under low nitrogen stress. Proportional phenotypic variance for the QTLs ranged from 0.7% to 15.22%, with an average of 5.05%.

Population 4 had the highest number of QTLs (36). This is a testament to the diversity within the studied F₃ populations across nitrogen management options. AD (26) and EH (24) had the highest number of QTLs whilst ASI (11) had the lowest. This is consistent with another study (Ertiro et al., 2020c) which found the highest QTL number as those underlying AD. In our study, this was consistent with the observed high genetic variance for AD and EH across the two nitrogen regimes.

TABLE 8 Summary of detected QTL for measured traits under optimum and low-nitrogen stress in F3 populations derived from seven elite lines.

Traits	Population	Management	No of QTL	TPVE* (%)
AD	F3pop1CML494×CML550	Optimum	1	8.09
	F3pop2CKL5017×CML536		9	18.89
	F3pop3CML550×CML507		3	11.80
	F3pop4VL081452×VL058589		1	12.64
	F3pop1CML494×CML550	Low N	1	7.70
	F3pop2CKL5017×CML536		9	18.55
	F3pop3CML550×CML507		1	7.45
	F3pop4VL081452×VL058589		4	35.24
ASI	F3pop1CML494×CML550	Low N	3	6.91
	F3pop4VL081452×VL058589		1	16.73
EH	F3pop1CML494×CML550	Optimum	4	23.20
	F3pop2CKL5017×CML536		4	26.35
	F3pop3CML550×CML507		3	22.25
	F3pop4VL081452×VL058589		8	15.26
	F3pop1CML494×CML550	Low N	1	4.87
	F3pop2CKL5017×CML536		4	14.42
	F3pop3CML550×CML507		4	13.43
EPO	F3pop1CML494×CML550	Optimum	2	15.56
	F3pop2CKL5017×CML536		2	11.13
	F3pop1CML494×CML550	Low N	2	11.56
	F3pop2CKL5017×CML536		2	17.98
GY	F3pop1CML494×CML550	Optimum	2	11.50
	F3pop2CKL5017×CML536		2	16.52
	F3pop3CML550×CML507		2	12.15
	F3pop4VL081452×VL058589		2	7.68
	F3pop2CKL5017×CML536	Low N	1	6.20
	F3pop3CML550×CML507		2	5.12
	F3pop4VL081452×VL058589		1	19.16
PH	F3pop1CML494×CML550	Optimum	1	6.01
	F3pop2CKL5017×CML536		4	25.07
	F3pop3CML550×CML507		5	25.81
PH	F3pop1CML494×CML550	Low N	3	9.11
	F3pop2CKL5017×CML536		4	18.76
	F3pop3CML550×CML507		1	7.27
	F3pop4VL081452×VL058589		1	6.34

*TPVE, total phenotypic variance explained; Low N- low soil N stress management; GY, Grain yield; AD, anthesis date; ASI, anthesis silking interval; PH, plant height; EH, ear height; EPO, ear position.

Marker-assisted selection for GY improvement relies on the successful identification of QTLs with moderate to high effects. For GY, 11 QTLs were detected under both optimum (7) and

low (4) nitrogen conditions. This is inconsistent with [Ertiro et al., 2020c](#) findings which recorded more QTLs under low nitrogen management than under optimum conditions. One

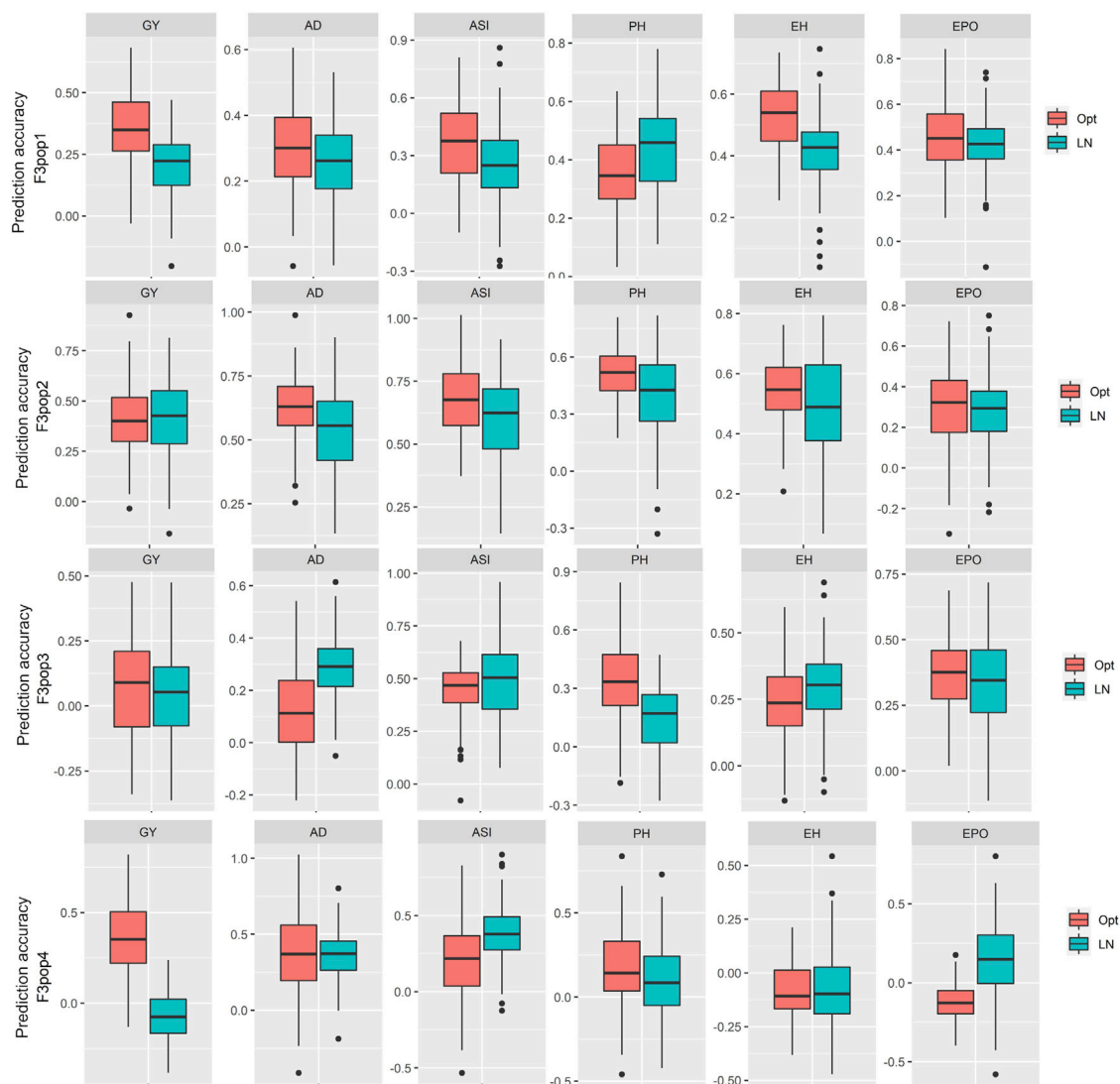


FIGURE 3

Distribution of the five-fold cross-validated genomic prediction correlations for grain yield (GY), anthesis date (AD), anthesis-silking interval (ASI), plant height (PH), ear height (EH), and ear position (EPO) evaluated in multiple environments under optimum and low soil N management conditions in four F_3 populations.

QTL for GY (*qGY10_130*) was found under optimum conditions on chromosome 10 in both population 3 (125.47–131.38 Mbp) and population 4 (46.11–130.61 Mbp). A study by Ertiro et al., 2020c also found QTLs underlying GY under different nitrogen regimes on chromosome 10. No common QTL for GY was identified across the studied F_3 populations. Ertiro et al. (2020a) also observed similar results, i.e., no common QTLs observed for GY. Noteworthy, the use of a different number of populations, locations and markers in future studies can give rise to a different outcome. In this study, some common QTLs were identified for AD, EH and EPO. The identification of common QTLs justifies the magnitude of correlation across the two nitrogen regimes. Common QTLs can also be used for indirect selection. Similar to the results of Ertiro et al., 2020c, QTL correspondences detected between the nitrogen regimes for the studied traits were not similar across F_3 populations and

this points to the genetic-background-specific nature of these QTLs.

The distinct genetic controls of the expression of the phenotypes reported in this study and related studies (Bänziger et al., 1999; Ribaut et al., 2007) suggest that selection for low nitrogen stress tolerance will be more effective under low nitrogen management. These findings unravel the most promising genomic regions for MAS to increase maize tolerance to low nitrogen stress. Further multi-environment trials with larger population sizes are necessary to validate the stability of the recorded QTLs and utilize them to identify underlying causal genes.

To achieve higher genetic gain, complex traits in maize breeding can be improved by integrating genomic, bioinformatic, and statistical tools into breeding programs (Beyene et al., 2019; Beyene et al., 2021). Rapid technological advancements have made available cheaper genotyping tools, while rapid progress in

TABLE 9 The prediction accuracy for grain yield and other traits evaluated under optimum and low-nitrogen stress conditions for four F₃ populations derived from seven elite lines.

Traits	GY	AD	ASI	PH	EH	EPO
Optimum						
F3pop1CML494×CML550	0.35	0.31	0.37	0.35	0.53	0.45
F3pop2CKL5017×CML536	0.41	0.63	0.68	0.52	0.54	0.29
F3pop3CML550×CML507	0.06	0.12	0.44	0.34	0.24	0.37
F3pop4VL081452×VL058589	0.35	0.37	0.20	0.18	−0.09	−0.11
Low N						
F3pop1CML494×CML550	0.20	0.26	0.25	0.44	0.42	0.42
F3pop2CKL5017×CML536	0.41	0.53	0.59	0.40	0.49	0.28
F3pop3CML550×CML507	0.04	0.29	0.50	0.15	0.29	0.35
F3pop4VL081452×VL058589	−0.07	0.36	0.38	0.09	−0.07	0.14

Low N- low soil N stress management; GY, Grain yield; AD, anthesis date; ASI, anthesis silking interval; PH, plant height; EH, ear height; EPO, ear position.

the field of big data and bioinformatics has led to the development of user-friendly software tools which can handle complex statistical models and facilitates breeders to apply them to improve complex traits on a more routine basis. GS, which predicts the breeding values by employing genome-wide markers is proving to be effective in improving complex traits that are controlled by multiple minor effect genes. For instance, RR-BLUP or G-BLUP is now very commonly used to predict several complex traits in maize (Guo et al., 2012; Hao et al., 2019; Beyene et al., 2021). Several NUE traits like chlorophyll index, chlorophyll fluorescence and leaf N content were also selected by genomic prediction before integrating into NUE breeding programs in ryegrass (Zhao et al., 2020). In the present study, GS was performed on GY and other agronomic traits evaluated under optimum and low soil N stress conditions. The average genomic prediction correlations ranged from 0.07 to 0.41 under optimum and −0.07 to 0.41 under low soil N stress conditions. The higher values of prediction correlations are comparable to earlier studies with a diversity panel evaluated under optimum and low soil N stress conditions (Ertiro et al., 2020c). The observed prediction accuracies for other agronomic traits are comparable to earlier studies reported under different stresses in maize (Zhang et al., 2015; Yuan et al., 2019).

In population 3, we found low accuracy for GY under optimum condition, which could be due to its small range of variability within the population as well as low heritability (Table 2, Figure 1). We observed a wide range of prediction correlations for same trait among different populations and management. This could be due to their differences in sample size, genetic variance, trait heritability, changes in population structure and linkage disequilibrium estimates. In some populations, we observed negative prediction correlations like GY for pop 4 under low soil N stress conditions (Figure 3; Table 9). Opposite linkage phases between markers and major-effect QTLs in the population may be the reason for negative correlations. In addition, the four F₃ populations used in this study were developed by using tolerant × tolerant crosses for low soil N stress, which may be another reason why we observed low prediction correlations (as most of the causal factors for these

traits might have been fixed on both the parental lines). As a result, variation is low for the traits which is evident with low to moderate heritability estimates (Table 2). Nevertheless, the prediction correlation followed a consistent pattern for all traits under both optimum and low N conditions. Traits with high prediction correlations also tend to have relatively high heritability estimates. In GS, the less complex trait AD and PH had higher accuracy compared to GY, which is consistent with the nature of trait complexity (Zhang et al., 2017a; Yuan et al., 2019). For simple inherited traits that are positively correlated with GY, prediction correlations are moderate to high, which clearly supports the usefulness of GS for their improvement under either optimum or low soil N stress conditions.

Overall, the linkage mapping studies with our F₃ populations revealed several QTL with minor to moderate effects for GY and other agronomic traits under optimum and low N management. It is difficult to capture multiple QTL with small to moderate effects for selection and their environmental and genotypic specific expressions make it even more difficult to improve these traits only through traditional breeding or a few QTL based MAS. However, discovery of genomic regions through linkage mapping will continue to be vital to understanding the genetic basis of these traits. On the other hand, genome-wide selection is critical in improving quantitative traits. The phenotypic selection efficiency per cycle is measured as h (square root of heritability) and the value of marker-based prediction accuracy close to h indicates the selection response based on markers and based on phenotypes are near equal (Dekkers, 2007; Lorenzana and Bernardo, 2009). The prediction accuracies observed for GY and other traits was $\geq 1/2h$ indicating that the response to marker-based selection would be at least 50% of the response to phenotypic selection for agronomic traits. With the possibility of making three cycles per year by using marker-based selection, the selection response will be −1.5 times the gain from one cycle of phenotypic selection. These results indicate that genome-wide selection would be more efficient in terms of genetic gain per year. However, one must be cautious as we observed prediction accuracy of <0.10 for some traits in some populations, where GS has no additional advantage over phenotypic selection.

5 Conclusion

Nitrogen-depleted soils are a major factor behind low maize productivity in smallholder farming systems in SSA. Genomics-based plant breeding techniques such as QTL mapping and GS can provide useful information for scaling the development of low nitrogen tolerant maize varieties. Here, we sought to identify the genomic regions associated with GY and related traits under optimum and low soil nitrogen conditions in F_3 maize populations grown in Kenya and Zimbabwe. Our analysis found a total of 91 QTLs underlying GY, ASI, AD, EH, EPO, and PH. However, no common QTL for GY was identified across the studied nitrogen regimes. The validation of these QTLs in the same F_3 populations is needed to guarantee the success of future marker-assisted selections. Identification of many QTL with minor effects indicates rather QTL mapping on these traits, GS is more appropriate for their improvement. Genomic prediction correlations were low to moderate. However, by considering the possibility to have three cycles per year with marker-based selection, integration of GS in a low N tolerant breeding program will be more efficient to improve GY under low soil N conditions.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. This data can also be accessed here: <https://zenodo.org/records/10021760>.

Author contributions

CK: Data curation, Formal Analysis, Investigation, Visualization, Writing–original draft. NN: Software, Validation, Visualization, Writing–original draft, Writing–review and editing. VC: Data curation, Investigation, Project administration, Writing–review and editing. BT: Investigation, Methodology, Visualization, Writing–review and editing. BD: Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Visualization, Writing–review and editing. YB: Validation, Visualization, Writing–review and editing. OK: Project administration, Resources, Supervision, Writing–review and editing. CS: Supervision, Validation, Writing–review and editing. BP: Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing–review and editing. MG: Conceptualization, Formal Analysis, Investigation, Methodology, Project administration, Software, Supervision, Validation, Visualization, Writing–original draft, Writing–review and editing.

References

- Agrama, H., Zakaria, A., Said, F., and Tuinstra, M. (1999). Identification of quantitative trait loci for nitrogen use efficiency in maize. *Mol. Breed.* 5, 187–195. doi:10.1023/a:1009669507144
- Bänziger, M., Bétrán, F., and Lafitte, H. (1997). Efficiency of high-nitrogen selection environments for improving maize for low-nitrogen target environments. *Crop Sci.* 37, 1103–1109. doi:10.2135/cropsci1997.0011183x003700040012x
- Bänziger, M., Edmeades, G., and Lafitte, H. (1999). Selection for drought tolerance increases maize yields across a range of nitrogen levels. *Crop Sci.* 39, 1035–1040. doi:10.2135/cropsci1999.0011183x003900040012x
- Baudron, F., Zaman-Allah, M. A., Chaipa, I., Chari, N., and Chinwada, P. (2019). Understanding the factors influencing fall armyworm (*Spodoptera frugiperda* J.E. Smith) damage in African smallholder maize fields and quantifying its impact on

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. The research was supported by the Bill and Melinda Gates Foundation (B&MGF), Foundation for Food and Agriculture Research (FFAR) and the United States Agency for International Development (USAID) through the Stress Tolerant Maize for Africa (STMA, B&MGF Grant # OPP1134248) Project, IMAS Project, AGG (Accelerating Genetic Gains in Maize and Wheat for Improved Livelihoods; B&MGF Investment ID INV-003439) Project and the CGIAR Research Program on Maize (MAIZE). NN and CS acknowledge funding from the European Commission DESIRA funded LEG4DEV Project [FOOD/2020/418-901] and the Science Foundation Ireland Principal Investigator Grant [13/IA/1820].

Acknowledgments

The authors thank the CIMMYT field technicians for phenotypic evaluations and Laboratory technicians for sample preparation for genotyping. We are also grateful to LGC genomic labs in the UK for genotyping the populations and providing the marker information. More gratitude to CIMMYT scientists for their excellent and noble guide throughout the processes of 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.

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/fgene.2023.1266402/full#supplementary-material>

yield. A case study in Eastern Zimbabwe. *A case study East. Zimbabwe Crop Prot.* 120, 141–150. doi:10.1016/j.cpropro.2019.01.028

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, 1502. 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, 685488. doi:10.3389/fpls.2021.685488

Beyene, Y., Semagn, K., Mugo, S., Tarekne, A., Babu, R., Meisel, B., et al. (2015). Genetic gains in grain yield through genomic selection in eight bi-parental maize populations under drought stress. *Crop Sci.* 55, 154–163. doi:10.2135/cropsci2014.07.0460

Boddupalli, P., Suresh, L., 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.virusres.2020.197943

Bonnett, D., Li, Y., Crossa, J., Dreisigacker, S., Basnet, B., Pérez-Rodríguez, P., et al. (2022). Response to early generation genomic selection for yield in wheat. *Front. Plant Sci.* 12, 718611. doi:10.3389/fpls.2021.718611

Chen, T., Liu, L., Zhou, Y., Zheng, Q., Luo, S., Xiang, T., et al. (2023a). Characterization and comprehensive evaluation of phenotypic characters in wild *Camellia oleifera* germplasm for conservation and breeding. *Front. Plant Sci.* 14, 1052890. doi:10.3389/fpls.2023.1052890

Chen, Y., Xiong, Y., Hong, H., Li, G., Gao, J., Guo, Q., et al. (2023b). Genetic dissection of and genomic selection for seed weight, pod length, and pod width in soybean. *Crop J.* 11, 832–841. doi:10.1016/j.cj.2022.11.006

CIMMYT (2005). *Cimmyt*. Mexico: DF.

Coque, M., and Gallais, A. (2006). Genomic regions involved in response to grain yield selection at high and low nitrogen fertilization in maize. *Theor. Appl. Genet.* 112, 1205–1220. doi:10.1007/s00122-006-0222-5

Crossa, J., Pérez-Rodríguez, P., Cuevas, J., Montesinos-López, O., Jarquín, D., De Los Campos, G., et al. (2017). Genomic selection in plant breeding: methods, models, and perspectives. *Trends plant Sci.* 22, 961–975. doi:10.1016/j.tplants.2017.08.011

Das, B., Atlin, G. N., Olsen, M., Burgueño, J., Tarekne, A., Babu, R., et al. (2019). Identification of donors for low-nitrogen stress with maize lethal necrosis (MLN) tolerance for maize breeding in sub-Saharan Africa. *Euphytica* 215, 1–15. doi:10.1007/s10681-019-2406-5

de Carvalho, E. V., Afférr, F. S., Peluzio, J. M., Dotto, M. A., and Cancellier, L. L. (2012). Nitrogen use efficiency in corn (*Zea mays* L.) genotypes under different conditions of nitrogen and seeding date. *Maydica* 57, 43–48.

De Groote, H., Kimenju, S. C., Munyua, B., Palmas, S., Kassie, M., and Bruce, A. (2020). Spread and impact of fall armyworm (*Spodoptera frugiperda* JE Smith) in maize production areas of Kenya. *Agric. Ecosyst. Environ.* 292, 106804. doi:10.1016/j.agee.2019.106804

Dekkers, J. C. M. (2007). Prediction of response to marker-assisted and genomic selection using selection index theory. *J. Animal Breed. Genet.* 124, 331–341. doi:10.1111/j.1439-0388.2007.00701.x

Derera, J., Tongona, P., Vivek, B. S., and Laing, M. D. (2008). Gene action controlling grain yield and secondary traits in southern African maize hybrids under drought and non-drought environments. *Euphytica* 162, 411–422. doi:10.1007/s10681-007-9582-4

Dreisigacker, S., Crossa, J., Pérez-Rodríguez, P., Montesinos-López, O., Rosyara, U., Juliana, P., et al. (2021). *Implementation of genomic selection in the CIMMYT global wheat program, findings from the past 10 years*. United Kingdom: Haptes.

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. *Glob. Food Secur.* 17, 48–56. doi:10.1016/j.gfs.2018.03.007

Eriksson, D., Akoroda, M., Azmach, G., Labuschagne, M., Mahungu, N., and Ortiz, R. (2018). Measuring the impact of plant breeding on sub-Saharan African staple crops. *Outlook Agric.* 47, 163–180. doi:10.1177/0030727018800723

Ertiro, B. T., Das, B., Kosgei, T., Tesfaye, A. T., Labuschagne, M. T., Worku, M., et al. (2022). Relationship between grain yield and quality traits under optimum and low-nitrogen stress environments in tropical maize. *Agronomy* 12, 438. doi:10.3390/agronomy12020438

Ertiro, B. T., Labuschagne, M., Olsen, M., Das, B., Prasanna, B. M., and Gowda, M. (2020a). Genetic dissection of nitrogen use efficiency in tropical maize through genome-wide association and genomic prediction. *Front. plant Sci.* 11, 474. doi:10.3389/fpls.2020.00474

Ertiro, B. T., Olsen, M., Das, B., Gowda, M., and Labuschagne, M. (2020b). Efficiency of indirect selection for grain yield in maize (*Zea mays* L.) under low nitrogen conditions through secondary traits under low nitrogen and grain yield under optimum conditions. *Euphytica* 216, 134–212. doi:10.1007/s10681-020-02668-w

Ertiro, B. T., Olsen, M., Das, B., Gowda, M., and Labuschagne, M. (2020c). Genetic dissection of grain yield and agronomic traits in maize under optimum and low-nitrogen stressed environments. *Int. J. Mol. Sci.* 21, 543. doi:10.3390/ijms21020543

Falconnier, G. N., Corbeels, M., Boote, K. J., Affholder, F., Adam, M., MacCarthy, D. S., et al. (2020). Modelling climate change impacts on maize yields under low nitrogen input conditions in sub-Saharan Africa. *Glob. change Biol.* 26, 5942–5964. doi:10.1111/gcb.15261

Ficht, A., Konkin, D. J., Cram, D., Sidebottom, C., Tan, Y., Pozniak, C., et al. (2023). Genomic selection for agronomic traits in a winter wheat breeding program. *Theor. Appl. Genet.* 136, 38. doi:10.1007/s00122-023-04294-1

Ganal, M. W., Durstewitz, G., Polley, A., Bérard, A., Buckler, E. S., Charcosset, A., et al. (2011). A large maize (*Zea mays* L.) SNP genotyping array: development and germplasm genotyping, and genetic mapping to compare with the B73 reference genome. *PLoS one* 6, e28334. doi:10.1371/journal.pone.0028334

García-Ruiz, A., Cole, J. B., VanRaden, P. M., Wiggins, G. R., Ruiz-López, F. J., and Van Tassel, C. P. (2016). Changes in genetic selection differentials and generation intervals in US Holstein dairy cattle as a result of genomic selection. *Proc. Natl. Acad. Sci.* 113, E3995–E4004. doi:10.1073/pnas.1519061113

Gowda, M., Makumbi, D., Das, B., Nyaga, C., Kosgei, T., Crossa, J., et al. (2021). Genetic dissection of *Striga hermonthica* (Del.) Benth. resistance via genome-wide association and genomic prediction in tropical maize germplasm. *Theor. Appl. Genet.* 134, 941–958. doi:10.1007/s00122-020-03744-4

Guo, Z., Tucker, D. M., Lu, J., Kishore, V., and Gay, G. (2012). Evaluation of genome-wide selection efficiency in maize nested association mapping populations. *Theor. Appl. Genet.* 124, 261–275. doi:10.1007/s00122-011-1702-9

Hao, Y., Wang, H., Yang, X., Zhang, H., He, C., Li, D., et al. (2019). Genomic prediction using existing historical data contributing to selection in biparental populations: a study of kernel oil in maize. *plant genome* 12, 180025. doi:10.3835/plantgenome2018.05.0025

Hirel, B., Le Gouis, J., Ney, B., and Gallais, A. (2007). The challenge of improving nitrogen use efficiency in crop plants: towards a more central role for genetic variability and quantitative genetics within integrated approaches. *J. Exp. Bot.* 58, 2369–2387. doi:10.1093/jxb/erm097

Hu, J., Chen, B., Zhao, J., Zhang, F., Xie, T., Xu, K., et al. (2022). Genomic selection and genetic architecture of agronomic traits during modern rapeseed breeding. *Nat. Genet.* 54, 694–704. doi:10.1038/s41588-022-01055-6

Juliana, P., He, X., Marza, F., Islam, R., Anwar, B., Poland, J., et al. (2022). Genomic selection for wheat blast in a diversity panel, breeding panel and full-sibs panel. *Front. Plant Sci.* 12, 745379. doi:10.3389/fpls.2021.745379

Kanampiu, F., Makumbi, D., Mageto, E., Omany, G., Waruingi, S., Musyoka, P., et al. (2018). Assessment of management options on *Striga* infestation and maize grain yield in Kenya. *Weed Sci.* 66, 516–524. doi:10.1017/wsc.2018.4

Kibe, M., Nyaga, C., Nair, S. K., Beyene, Y., Das, B., Bright, J. M., et al. (2020). 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

Kosambi, D. D. (2016). “The estimation of map distances from recombination values,” in *DD Kosambi* (Berlin, Germany: Springer), 125–130.

Liu, X.-H., He, S.-L., Zheng, Z.-P., Huang, Y.-B., Tan, Z.-B., Li, Z., et al. (2010). Identification of the QTLs for grain yield using RIL population under different nitrogen regimes in maize. *Afr. J. Agric. Res.* 5, 2002–2007. doi:10.1111/j.1439-0523.2007.01465.x

Lorenzana, R. E., and Bernardo, R. (2009). Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. *Theor. Appl. Genet.* 120, 151–161. doi:10.1007/s00122-009-1166-3

Meng, L., Li, H., Zhang, L., and Wang, J. (2015). QTL IciMapping: integrated software for genetic linkage map construction and quantitative trait locus mapping in biparental populations. *Crop J.* 3 (3), 269–283. doi:10.1016/j.cj.2015.01.001

Meuwissen, T. H., Hayes, B. J., and Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, 1819–1829. doi:10.1093/genetics/157.4.1819

Ndlovu, N. (2020). Application of genomics and phenomics in plant breeding for climate resilience. *Asian Plant Res. J.*, 53–66. doi:10.9734/aprj/2020/v6i430137

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

Qin, J., Wang, F., Zhao, Q., Shi, A., Zhao, T., Song, Q., et al. (2022). Identification of candidate genes and genomic selection for seed protein in soybean breeding pipeline. *Front. Plant Sci.* 13, 882732. doi:10.3389/fpls.2022.882732

Ribaut, J.-M., Fracheboud, Y., Monneveux, P., Banziger, M., Vargas, M., and Jiang, C. (2007). Quantitative trait loci for yield and correlated traits under high and low soil nitrogen conditions in tropical maize. *Mol. Breed.* 20, 15–29. doi:10.1007/s11032-006-9041-2

Ribaut, J.-M., Jiang, C., Gonzalez-de-Leon, D., Edmeades, G., and Hoisington, D. (1997). Identification of quantitative trait loci under drought conditions in tropical maize. 2. yield components and marker-assisted selection strategies: 2. yield components and marker-assisted selection strategies. *Theor. Appl. Genet.* 94, 887–896. doi:10.1007/s001220050492

- Semagn, K., Beyene, Y., Babu, R., Nair, S., Gowda, M., Das, B., et al. (2015). Quantitative trait loci mapping and molecular breeding for developing stress resilient maize for sub-saharan africa. *Crop Sci.* 55, 1449–1459. doi:10.2135/cropsci2014.09.0646
- Semagn, K., Bjørnstad, Å., and Xu, Y. (2010). The genetic dissection of quantitative traits in crops. *Electron. J. Biotechnol.* 13, 0–17. doi:10.2225/vol13-issue5-fulltext-14
- Semagn, K., Magorokosho, C., Vivek, B. S., Makumbi, D., Beyene, Y., Mugo, S., et al. (2012). Molecular characterization of diverse CIMMYT maize inbred lines from eastern and southern Africa using single nucleotide polymorphic markers. *BMC genomics* 13, 113–211. doi:10.1186/1471-2164-13-113
- Shitta, N. S., Abtew, W. G., Ndlovu, N., Oselebe, H. O., Edemodu, A. C., and Abebe, A. T. (2021). Morphological characterization and genotypic identity of African yam bean (*Sphenostylis stenocarpa* Hochst ex. A. Rich. Harms) germplasm from diverse ecological zones. *Plant Genet. Resour.* 19, 58–66. doi:10.1017/s1479262121000095
- Tesfaye, K., Gbегbelegbe, S., Cairns, J. E., Shiferaw, B., Prasanna, B. M., Sonder, K., et al. (2015). Maize systems under climate change in sub-Saharan Africa: potential impacts on production and food security. *Int. J. Clim. Change Strategies Manag.* 7, 247–271. doi:10.1108/ijccsm-01-2014-0005
- Vivek, B. S., Krishna, G. K., Vengadessan, V., Babu, R., Zaidi, P. H., Kha, L. Q., et al. (2017). Use of genomic estimated breeding values results in rapid genetic gains for drought tolerance in maize. *Plant Genome* 10. doi:10.3835/plantgenome2016.07.0070
- Wang, N., Wang, H., Zhang, A., Liu, Y., Yu, D., Hao, Z., et al. (2020). Genomic prediction across years in a maize doubled haploid breeding program to accelerate early-stage testcross testing. *Theor. Appl. Genet.* 133, 2869–2879. doi:10.1007/s00122-020-03638-5
- Worku, M., Banziger, M., Erley, G., Friesen, D., Diallo, A. O., and Horst, W. J. (2012). Nitrogen efficiency as related to dry matter partitioning and root system size in tropical mid-altitude maize hybrids under different levels of nitrogen stress. *Field Crops Res.* 130, 57–67. doi:10.1016/j.fcr.2012.02.015
- Worku, M., Banziger, M., Erley, G. S. A., Friesen, D., Diallo, A. O., and Horst, W. J. (2007). Nitrogen uptake and utilization in contrasting nitrogen efficient tropical maize hybrids. *Crop Sci.* 47, 519–528. doi:10.2135/cropsci2005.05.0070
- Worku, M., Banziger, M., Friesen, D., Schulte, G., Horst, W. J., and Vivek, B. S. (2008). Relative importance of general combining ability and specific combining ability among tropical maize (*Zea mays* L.) inbreds under contrasting nitrogen environments. *Maydica* 53, 279–288.
- Würschum, T. (2012). Mapping QTL for agronomic traits in breeding populations. *Theor. Appl. Genet.* 125, 201–210. doi:10.1007/s00122-012-1887-6
- Yacoubou, A. M., Zoumarou Wallis, N., Menkir, A., Zinsou, V. A., Onzo, A., Garcia-Oliveira, A. L., et al. (2021). Breeding maize (*Zea mays*) for Striga resistance: past, current and prospects in sub-saharan africa. *Plant Breed.* 140, 195–210. doi:10.1111/pbr.12896
- Yuan, Y., Cairns, J. E., Babu, R., Gowda, M., Makumbi, D., Magorokosho, C., et al. (2019). Genome-wide association mapping and genomic prediction analyses reveal the genetic architecture of grain yield and flowering time under drought and heat stress conditions in maize. *Front. plant Sci.* 9, 1919. doi:10.3389/fpls.2018.01919
- Zaman-Allah, M., Das, B., Cairns, J. E., Vinayan, M. T., Tarekgegne, A., Magorokosho, C., et al. (2018). Phenotyping for abiotic stress tolerance in maize—low nitrogen stress. A field manual; CIMMYT: harare, Zimbabwe. Available online: <https://repository.cimmyt.org/bitstream/handle/10883/19935/60069.pdf?sequence=1> (accessed on June 10, 2023).
- Zhang, A., Wang, H., Beyene, Y., Semagn, K., Liu, Y., Cao, S., et al. (2017a). Effect of trait heritability, training population size and marker density on genomic prediction accuracy estimation in 22 bi-parental tropical maize populations. *Front. Plant Sci.* 8, 1916. doi:10.3389/fpls.2017.01916
- Zhang, X., Pérez-Rodríguez, P., Burgueño, J., Olsen, M., Buckler, E., Atlin, G., et al. (2017b). Rapid cycling genomic selection in a multiparental tropical maize population. *G3 Genes, Genomes, Genet.* 7, 2315–2326. doi:10.1534/g3.117.043141
- Zhang, X., Pérez-Rodríguez, P., Semagn, K., Beyene, Y., Babu, R., López-Cruz, M., et al. (2015). Genomic prediction in biparental tropical maize populations in water-stressed and well-watered environments using low-density and GBS SNPs. *Heredity* 114, 291–299. doi:10.1038/hdy.2014.99
- Zhao, X., Nie, G., Yao, Y., Ji, Z., Gao, J., Wang, X., et al. (2020). Natural variation and genomic prediction of growth, physiological traits, and nitrogen-use efficiency in perennial ryegrass under low-nitrogen stress. *J. Exp. Bot.* 71, 6670–6683. doi:10.1093/jxb/eraa388
- 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



OPEN ACCESS

EDITED BY

Amaranatha Reddy Vennapusa,
Delaware State University, United States

REVIEWED BY

Aamir W. Khan,
University of Missouri, United States
Aarthi Thiagaraya Selvam,
Texas A&M University, United States
Shailesh Sharma,
National Institute of Animal
Biotechnology (NIAB), India

*CORRESPONDENCE

Xiaochuan Sun,
✉ xchsun1987@163.com

[†]These authors have contributed equally
to this work

RECEIVED 31 May 2023

ACCEPTED 17 October 2023

PUBLISHED 02 November 2023

CITATION

Sun X, Tang M, Xu L, Luo X, Shang Y,
Duan W, Huang Z, Jin C and Chen G
(2023), Genome-wide identification of
long non-coding RNAs and their potential
functions in radish response to salt stress.
Front. Genet. 14:1232363.
doi: 10.3389/fgene.2023.1232363

COPYRIGHT

© 2023 Sun, Tang, Xu, Luo, Shang, Duan,
Huang, Jin and Chen. This is an open-
access article distributed under the terms
of the [Creative Commons Attribution
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). 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 of long non-coding RNAs and their potential functions in radish response to salt stress

Xiaochuan Sun^{1*†}, Mingjia Tang^{2†}, Liang Xu³, Xiaobo Luo⁴,
Yutong Shang⁴, Weihe Duan¹, Zhinan Huang¹, Cong Jin¹ and
Guodong Chen¹

¹School of Life Science and Food Engineering, Huaiyin Institute of Technology, Huaian, China,

²Department of Horticulture, Zhejiang University, Hangzhou, China, ³National Key Laboratory of Crop Genetics & Germplasm Enhancement and Utilization, Key Laboratory of Horticultural Crop Biology and Genetic Improvement (East China) of MOAR, College of Horticulture, Nanjing Agricultural University, Nanjing, China, ⁴Guizhou Institute of Biotechnology, Guizhou Province Academy of Agricultural Sciences, Guiyang, China

Long non-coding RNAs (lncRNAs) are increasingly recognized as cis- and trans-acting regulators of protein-coding genes in plants, particularly in response to abiotic stressors. Among these stressors, high soil salinity poses a significant challenge to crop productivity. Radish (*Raphanus sativus* L.) is a prominent root vegetable crop that exhibits moderate susceptibility to salt stress, particularly during the seedling stage. Nevertheless, the precise regulatory mechanisms through which lncRNAs contribute to salt response in radish remain largely unexplored. In this study, we performed genome-wide identification of lncRNAs using strand-specific RNA sequencing on radish fleshy root samples subjected to varying time points of salinity treatment. A total of 7,709 novel lncRNAs were identified, with 363 of them displaying significant differential expression in response to salt application. Furthermore, through target gene prediction, 5,006 cis- and 5,983 trans-target genes were obtained for the differentially expressed lncRNAs. The predicted target genes of these salt-responsive lncRNAs exhibited strong associations with various plant defense mechanisms, including signal perception and transduction, transcription regulation, ion homeostasis, osmoregulation, reactive oxygen species scavenging, photosynthesis, phytohormone regulation, and kinase activity. Notably, this study represents the first comprehensive genome-wide analysis of salt-responsive lncRNAs in radish, to the best of our knowledge. These findings provide a basis for future functional analysis of lncRNAs implicated in the defense response of radish against high salinity, which will aid in further understanding the regulatory mechanisms underlying radish response to salt stress.

KEYWORDS

lncRNA, salt stress, *Raphanus sativus* L., RNA-seq, target genes, Gene Ontology and pathway analyses

Introduction

Soil salinity is one of the major damaging issues to cultivable land, resulting in an inevitable slowdown to plant growth and yields (Safdar et al., 2019). It is estimated that approximately 20% of global irrigated land is affected by salinization, which negatively affects land productivity (Islam et al., 2022). Hyper-ionic and hyper-osmotic injuries, as well as oxidative damage caused by prolonged saline conditions, are the main effects of salt stress in plants. To a certain degree, plants have evolved adaptive mechanisms to ensure their survival under high salinity through molecular, biochemical, and physiological adjustments. Hence, it is imperative to unravel the molecular and physiological mechanisms associated with salt stress tolerance in order to address the substantial agronomic challenges posed by salinization.

Long non-coding RNAs (lncRNAs) refer to a type of non-coding RNAs (ncRNAs) greater than 200 nucleotides (nt) in length without discernible coding potential in eukaryotes (Kapranov et al., 2007; Wang and Chekanova, 2017). Similar to classical messenger RNAs (mRNAs), lncRNAs are primarily transcribed by RNA Pol II, capped, polyadenylated, and usually spliced, but generally exhibit lower expression levels (Chekanova, 2015; Quan et al., 2015). In plants, a subset of lncRNAs is also transcribed by two other plant-specific RNA polymerases, namely, Pol IV and V, and is usually synthesized from transposable elements (St Laurent et al., 2015; Wierzbicki et al., 2021). Based on their genomic location and cellular function, lncRNAs can be classified into sense and antisense lncRNAs, intronic lncRNAs, and intergenic lncRNAs (lincRNAs) (Ma et al., 2013). Substantial evidence supports the notion that lncRNAs play crucial roles as regulators in various biological processes, exerting their influence on gene expression in both cis- and trans-acting manners (Gil and Ulitsky, 2020; Jha et al., 2020); they can serve as molecular signals, decoys, guides, and scaffolds (Wang and Chang, 2011).

Initially, lncRNAs were neglected as a component of transcriptional noise (Chen et al., 2018); nevertheless, accumulating evidence has manifested that lncRNAs play crucial roles in multiple plant biological events, including photomorphogenesis (Wang et al., 2014; Sun et al., 2020), senescence (Huang et al., 2021), reproduction (Zhang et al., 2014; Zhou et al., 2022a), seed aging (Zhang et al., 2022), and fruit ripening (Tian et al., 2019; Zhou et al., 2022b). To date, there have been experimental investigations into the functions of specific plant lncRNAs. For instance, lncRNA39026 in tomato may possess the capability to impact decoy miR168a, leading to an increase in the expression of pathogenesis-related genes, thereby improving disease resistance (Hou et al., 2020). In rice, an intronic lncRNA known as RICE FLOWERING ASSOCIATED (RIFLA) plays an essential role in the flowering process mediated by *OsMADS56*, achieved through the formation of a complex with *OsIEZ1* (Shin et al., 2022). Recent research has hinted that lncRNAs exhibit responsiveness to various stressors, including salt stress (Baruah et al., 2021; Cao et al., 2021; Liu et al., 2022a; Hu et al., 2022). For example, a total of 742 salt-responsive lncRNAs were identified in maize (Liu et al., 2022a). Similarly, in *Spirodela polyrhiza*, a total of 2,815 lncRNAs were discovered under salt stress conditions, out of which 185 exhibited differential expression in response to salinity (Fu et al., 2020). In the context of cotton, the functional role of lncRNA973 in fine-tuning

salt stress response has been experimentally verified (Zhang et al., 2019). Additionally, the overexpression of lncRNA77580 in *Glycine max* has been found to regulate the response to both salt and drought stress by modulating the transcription of distinct sets of stress-related genes (Chen et al., 2023). Overall, these reports provide substantial evidence to affirm the significant role of lncRNAs in the regulation of plant stress tolerance.

Radish (*Raphanus sativus* L.), an essential root vegetable belonging to the Brassicaceae family, is cultivated globally owing to its high nutritional and medicinal values. Among vegetable crops, radish exhibits moderate sensitivity to salt, with its edible fleshy taproots displaying notable responsiveness to salt stimulus up to a maximum soil salinity threshold of 1.2 dS/m (Grattan et al., 2016). The yield and quality of radish taproots are significantly impacted by salt stress, as it has become the major limiting factor due to soil salinization and secondary salinization. Investigating the response of radish to salt stress holds potential for the advancement of salt-tolerant radish lines in the field of radish breeding. In radish, a substantial number of miRNAs and protein-coding genes implicated in salt adaptation have been reported in our previous studies (Sun et al., 2015; 2016). Nevertheless, the reports on lncRNAs involved in salt stress response remain elusive in radish. Fortunately, with the notable advancements in deep transcriptome sequencing technology and associated bioinformatics methodologies, it is now easier to comprehensively mine novel non-coding RNA molecules. In this study, strand-specific RNA sequencing (ssRNA-seq) was utilized to identify and characterize the salt-responsive lncRNAs in radish. Furthermore, the function of differentially expressed lncRNAs (DE-lncRNAs) was investigated by examining their position or co-expression associations with target genes. These findings could serve as a starting point for discerning the role of lncRNAs in the regulatory mechanisms governing salt stress response in radish, thereby offering potential benefits for the development of salt-tolerant radish cultivars.

Materials and methods

Plant materials and salt treatments

The radish advanced inbred line, 'YH', was used in this study. The seeds were rinsed and surface-sterilized using a 1.2% NaClO solution before germinating on moist filter paper and further incubated at 25°C in darkness for 2 days. Subsequently, the seedlings were transferred to plastic pots and grown at 25°C/18°C with a relative humidity of 60% in a 14 h light/10 h dark cycle. Uniform four-leaf-old seedlings were transferred into hydroponic conditions with half-strength Hoagland's solution. After acclimating for a 1-week slow seeding period, the plants were stressed with 200 mM NaCl. The materials for transcriptome sequencing were collected at 0 h (control), 6 h, 12 h, 24 h, 48 h, and 96 h under salt treatment with two biological replicates at each time point, and the sample for each replicate contained an equal amount of fleshy taproot from three individual seedlings. Additionally, the samples were prepared for qualitative real-time polymerase chain reaction (qRT-PCR) analysis in triplicate at every time point during salt treatments, with three individual seedlings per replicate. Last, the samples were frozen in liquid nitrogen immediately and stored at −80°C until use.

RNA extraction, quality control, and cDNA library preparation

The construction of transcriptome libraries and deep sequencing were implemented at the Novogene Bioinformatics Institute (Beijing, China). The isolation of total RNAs was performed following the manufacturer's recommended protocol (Invitrogen, Carlsbad, CA, United States). RNA quality and integrity were monitored using the Agilent 2100 Bioanalyzer. Ribosomal RNA was filtered using the Ribo-Zero™ rRNA Removal Kit (Epicentre, United States). The cDNA libraries were generated using the NEBNext® Ultra™ II RNA Kits (NEB, United States).

Sequencing and bioinformatic discovery of lncRNAs

The constructed libraries underwent sequencing using the Illumina NovaSeq PE150 platform. To obtain clean reads, the raw reads were filtered with fastp to eliminate adapter sequences and reads containing poly-N and low-quality reads. Meanwhile, the Q20, Q30, and GC contents were calculated. The clean reads were then aligned to the radish reference genome Rs1.0 assembly (<http://radish-genome.org/>) using HISAT2 (version 2.1.0) (Kim et al., 2015). Based on the assembled transcripts, the candidate lncRNAs were identified using a rigorous set of criteria. Initially, transcripts possessing a class_code of “i”, “u”, “x”, and “o” were reserved. Subsequently, transcripts exceeding a length of 200 bp and containing a minimum of two exons were retained. Finally, transcripts exhibiting overlap with established protein-coding domains were removed, while those overlapping with annotated lncRNAs were preserved. The coding potential of the retained transcripts was assessed using the Coding Potential Calculator (CPC2), Coding-Non-Coding-Index (CNCI), and Pfam. Non-coding transcripts shared by these three predictive software packages were identified as novel lncRNAs.

Differential expression analysis

The StringTie program (version 1.3.4) using the “-G option” (Pertea et al., 2015), in conjunction with the calculated FPKM values, was performed to estimate the expression level of both lncRNAs and coding genes in each sample. The DESeq2 R package (version 1.10.1) (Love et al., 2014) was applied to identify differentially expressed genes (DEGs) and DE-lncRNAs between the control and salinity treatments. Genes or lncRNAs with adjusted p -values (p_{adj}) ≤ 0.05 and \log_2 (|fold change|) ≥ 1 were determined as DEGs and DE-lncRNAs, respectively.

Potential target gene prediction and enrichment analysis of DE-lncRNAs

In order to enhance the understanding of lncRNA functions in radish, the potential cis- and trans-target mRNAs of DE-lncRNAs were predicted. According to the genomic location of lncRNAs relative to the neighboring genes, the target mRNAs in the 100-kb region upstream or downstream of DE-lncRNAs were regarded as potential cis-target genes. Pearson's correlation coefficient (PCC) between DE-lncRNAs

and the corresponding transcripts was calculated based on their co-expression to predict the trans-target genes of lncRNAs. The transcripts up to the strict standards of (|PCC| > 0.8 , p -value < 0.01) were defined as the trans-target genes of DE-lncRNAs. Subsequently, Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses were carried out to further annotate the functions of cis- and trans-target genes for DE-lncRNAs. The GO terms and KEGG pathways with corrected p -values ≤ 0.05 were recognized as significantly enriched.

Validation of DE-lncRNAs by qRT-PCR analysis

In order to validate the findings from RNA-seq analysis, a total of six DE-lncRNAs, namely, TCONS_00053136, TCONS_00062931, TCONS_00081369, TCONS_00101865, TCONS_00122106, and TCONS_00159796, were randomly selected for qRT-PCR analysis based on their significant differential expression observed in the RNA-seq data under salt stress conditions. The specific primers used for qRT-PCR analysis are listed in [Supplementary Table S1](#). The expression of the DE-lncRNAs was normalized using *RsActin* as the internal reference. qRT-PCRs were operated on a LightCycler R 480 System (Roche, Mannheim, and Germany) using the $2^{-\Delta\Delta C_T}$ method.

Results

Overview of whole-transcriptome sequencing in radish

To comprehensively identify the lncRNAs responsive to salt stress in radish, high-throughput ssRNA-seq of twelve libraries derived from samples of both the control (CK) and salinity treatment groups (NA_1-NA_5) were analyzed. The raw sequencing data have been deposited in the Sequence Read Archive (SRA) at the NCBI, with BioProject No. PRJNA930138. Following quality control measures on the raw data, a total of 1.11 billion clean reads (with an average of 92.71 million reads) were obtained. The Q20 and Q30 scores exceeded 90%, and the GC content ranged from 45.31% to 48.98%. All clean reads were aligned to the radish reference genome, with the alignment ratio ranging from 70.35% to 79.85%. The statistics of sequencing and mapping for each library are summarized in [Supplementary Table S2](#).

Identification of novel lncRNAs in radish

After transcriptome assembly using the StringTie software, a total of 708,647 assembled transcripts were obtained from the mapped reads of the twelve libraries. To predict the coding potential of these novel assembled transcripts, two prediction software programs (CPC2 and CNCI) and one database (Pfam) were employed. Consequently, 7,709 transcripts were identified as novel lncRNAs ([Supplementary Table S3](#)). Meanwhile, 8,288 transcripts were determined to be novel mRNAs ([Supplementary Table S4](#)). The lncRNAs were characterized based on their relative proximity to the nearest protein-coding

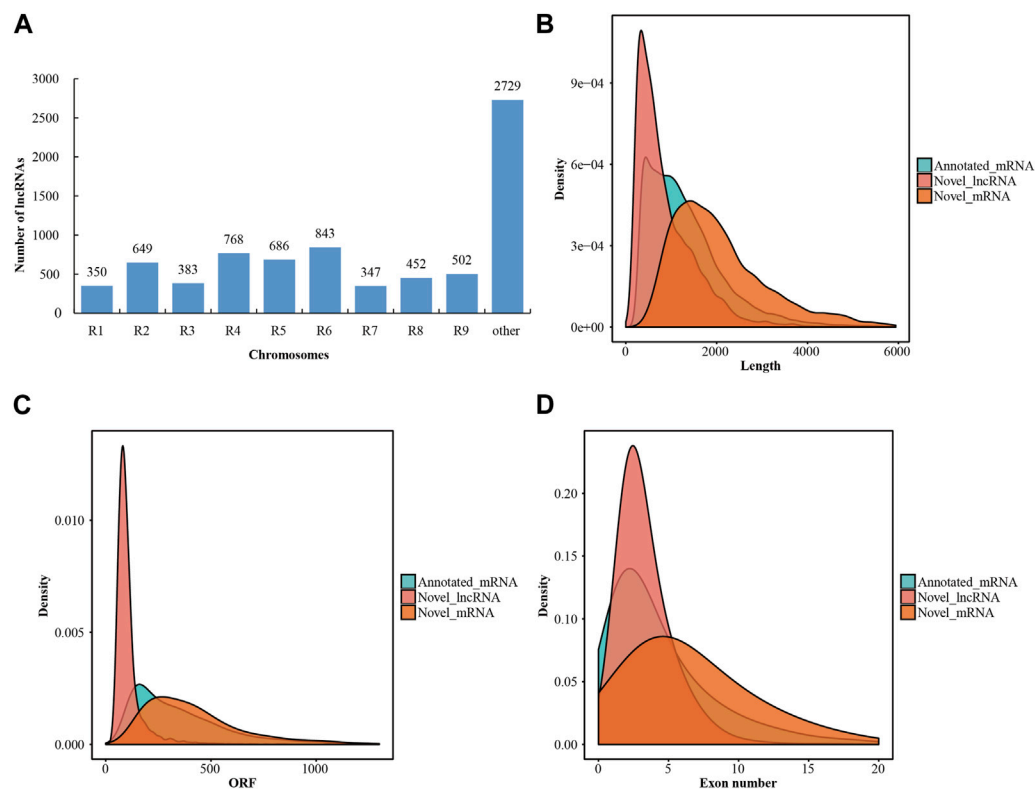


FIGURE 1

Characteristics of lncRNAs in radish. **(A)** Distribution of lncRNAs on different chromosomes. **(B)** Statistical results of length distribution between the lncRNAs and mRNAs. **(C)** Statistical results of ORF distribution between lncRNAs and mRNAs. **(D)** Statistical results of exon number percentages between lncRNAs and mRNAs.

gene neighbor, with all lncRNAs located as lincRNAs in the intergenic region. Approximately two-thirds of the lncRNAs were distributed across all the radish chromosomes, leaving more than one-third of the lncRNAs unmapped to any specific chromosome (Figure 1A; Supplementary Table S3). In addition, a comparative analysis was conducted to assess the fundamental characteristics of lncRNAs and mRNAs, encompassing sequence length, open reading frames (ORFs), and exon numbers. The length distribution of lncRNAs in radish predominantly fell within the range of 200–7,000 bp, with approximately 93.4% of lncRNAs falling within the 200–2,000 bp range (Figure 1B). Additionally, radish lncRNAs also harbored shorter ORFs compared to mRNAs, with the majority of novel lncRNAs having ORFs of less than 150 nucleotides (Figure 1C). Structural analysis further indicated that a significant proportion (67.4%) of novel lncRNAs possessed fewer than three exons, whereas mRNAs displayed a higher exon count (Figure 1D).

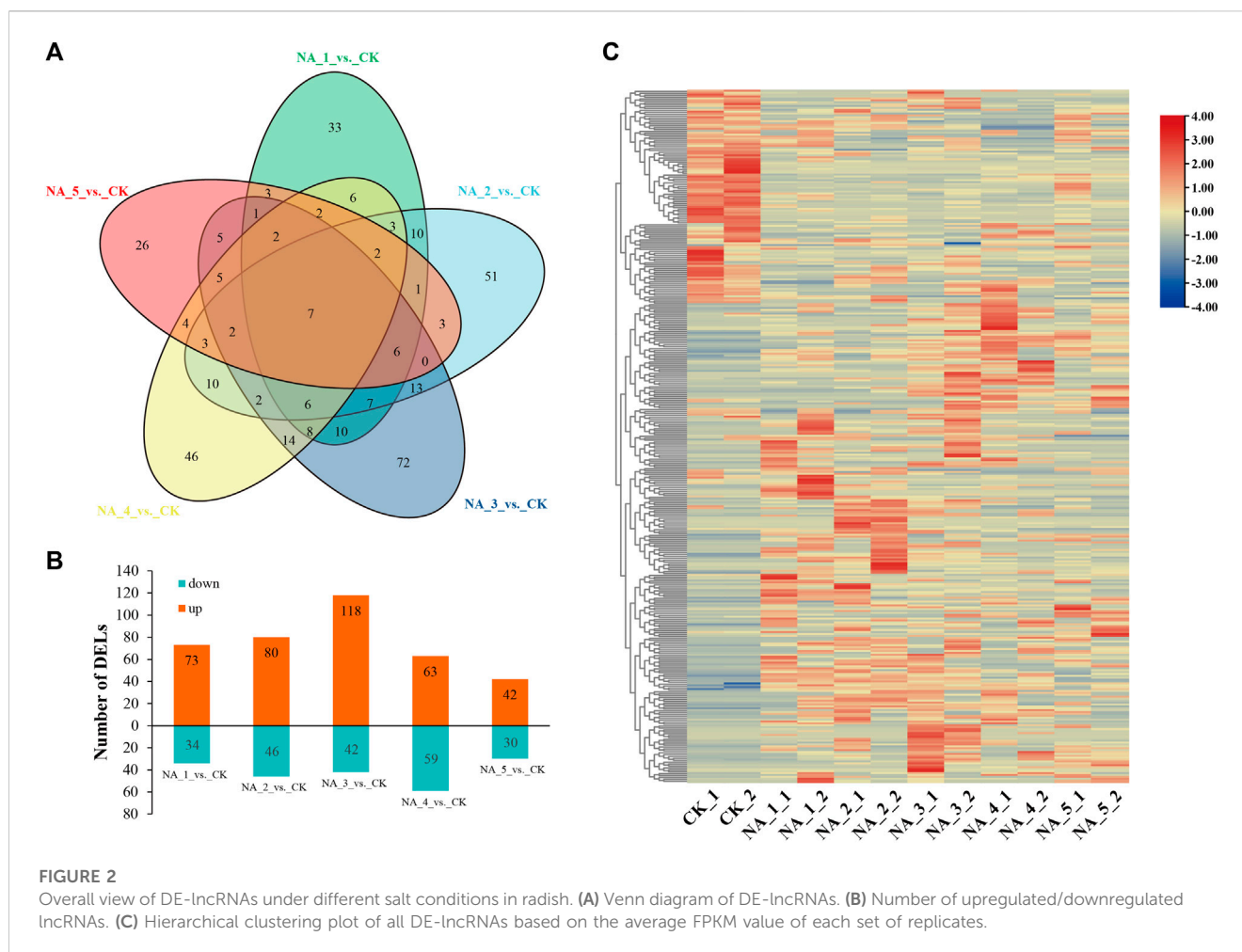
Identification and analysis of DE-lncRNAs in radish

In order to investigate the changes in transcription of lncRNAs during salt stress, we conducted a comprehensive analysis using the edgeR package (version 3.12.1). DE-lncRNAs were filtered based on the conditions $p_{adj} \leq 0.05$ and $\log_2(|\text{fold change}|) \geq 1$. In total, 363 DE-lncRNAs were identified, with 107, 126, 160, 122, and

72 members detected in the NA_1_vs._CK, NA_2_vs._CK, NA_3_vs._CK, NA_4_vs._CK, and NA_5_vs._CK comparisons, respectively (Supplementary Table S5). Only seven common DE-lncRNAs were traced in five comparisons (Figure 2A). In the comparison between NA_1_vs._CK, a total of 73 lncRNAs were found to be upregulated, while 34 lncRNAs were downregulated. Similarly, in the comparison between NA_2_vs._CK, 80 lncRNAs were identified as upregulated and 46 lncRNAs were downregulated. Furthermore, in the comparison between NA_3_vs._CK, a total of 118 lncRNAs were upregulated and 42 lncRNAs were downregulated. Additionally, in the comparison between NA_4_vs._CK, 63 lncRNAs were discovered to be upregulated and 59 lncRNAs were downregulated. Last, in the comparison between NA_5_vs._CK, 42 lncRNAs were found to be upregulated and 30 lncRNAs were downregulated. Consequently, the prevalence of upregulated lncRNAs was observed in all comparisons, as depicted in Figure 2B. Furthermore, a heat map was employed to aggregate the DE-lncRNAs, enabling the visualization of expression patterns across all six time points of salt treatments (Figure 2C).

Analysis of potential target genes of DE-lncRNAs in radish

To explore the potential biological roles of lncRNAs in response to salinity stressors in radish, the target genes fine-tuned by DE-



lncRNAs were identified using both cis- and trans-acting approaches (Supplementary Table S6). In terms of cis-regulation, a total of 5,006 target genes were observed for 303 DE-lncRNAs depending on their genomic location, resulting in 6,774 lncRNA-mRNA interaction pairs. Additionally, 5,983 trans-target genes were recognized for 249 DE-lncRNAs according to their co-expression relationships, resulting in 24,059 lncRNA-mRNA interaction pairs. Notably, 599 target genes were found to be jointly regulated by DE-lncRNAs via both cis- and trans-acting approaches. Moreover, the study found that 89 target genes (1.78%) exhibited differential expression under salt stress with 121 cis-regulatory pairs, and 229 target genes (3.83%) showed differential expression with 1,712 trans-regulatory pairs (Tables 1, 2; Supplementary Table S7). To further distinguish the potential key target genes responsive to salt stress, further annotation and analysis were conducted on these 121 cis-regulatory pairs and 1,712 trans-regulatory pairs. Among the differentially expressed targeted genes, there were crucial genes encoding for transcription factors (TFs) (e.g., bHLH, WRKY, MYB, and MADS-box), transport-related proteins (e.g., calcium-transporting ATPase, vacuolar glucose transporter, annexin, sugar transporter, and sulfate transporter), cytochrome P450, receptor-like protein kinases, enzymes associated with signal transduction (e.g., mitogen-activated protein kinase, calmodulin, CBL-interacting protein kinase, and CBL-interacting

serine/threonine-protein kinase), enzymes related to photosynthesis (e.g., water-soluble chlorophyll protein), plant hormone regulation (e.g., ACC oxidase, auxin-responsive GH3 family protein, and proline dehydrogenase), osmoregulatory factors like trehalose-6-phosphate phosphatase, and enzymes implicated in mitigating oxidative damage (e.g., superoxide dismutase, peroxidase, and thioredoxin). All in all, these findings hinted that the identified DE-lncRNA-target pairs may play essential roles in regulating the significant biological processes related to the response of radish to salt stress.

Enrichment analysis for cis-acting target genes of DE-lncRNAs

GO analysis exhibited that the cis-target genes of DE-lncRNAs were enriched in 2,961 GO terms, including 1,677 biological processes (BPs), 376 molecular functions (MFs), and 908 cellular components (CCs) (Supplementary Table S8). Increasingly, GO analysis of cis-acting targets responsive to salt stress was closely enriched in BP terms related to photosynthesis (e.g., GO: 0015979: photosynthesis and GO: 0015994: chlorophyll metabolic process), signal transduction (e.g., GO: 0009966: regulation of signal transduction; GO: 0023051: regulation of signaling; GO: 0009755: hormone-mediated signaling pathway; and GO: 0006465: signal

TABLE 1 Representative differentially expressed cis-targets for some salt-responsive lncRNAs.

lncRNA_ID	Gene_ID	Description of the target gene	Gene name	Distance	cis location
TCONS_00007404	Rs013430	Receptor-like protein kinase-like	—	14,586	Upstream
TCONS_00010534	Rs038050	TSK-associating protein 1	<i>TSA1</i>	13,490	Downstream
TCONS_00015424	Rs068580	Leucine-rich repeat (LRR) family protein	—	75,018	Downstream
TCONS_00020107	Rs069760	Protein TIFY 11b	—	87,973	Upstream
TCONS_00021725	Rs228720	LRR family protein FLR1	<i>FLR1</i>	94,115	Downstream
TCONS_00034882	XLOC_021821	Ubiquitin-related modifier 1 homolog 1	—	10,293	Downstream
TCONS_00035170	XLOC_018600	Mitogen-activated protein kinase 20-like	<i>MAPK20</i>	−32,989	Antisense
TCONS_00036718	Rs133050	Calcium-transporting ATPase 8	—	9,985	Downstream
TCONS_00049496	Rs159740	ACT domain repeat 7	<i>ACR7</i>	37,198	Downstream
	Rs159870	Trehalose-6-phosphate phosphatase	<i>TPPB</i>	41,249	Upstream
TCONS_00053221	XLOC_033420	Calmodulin-1	<i>CAM1</i>	−2,821	Sense
TCONS_00053239	XLOC_033420	Calmodulin-1	<i>CAM1</i>	82,458	Upstream
TCONS_00058325	Rs223580	Protein dehydration-induced 19-4	—	17,702	Upstream
TCONS_00060741	Rs240270	F-box kelch-repeat protein	—	74,856	Downstream
TCONS_00066609	XLOC_047803	Cytochrome P450 72A15-like	—	−2,858	Antisense
	XLOC_047805	Cytochrome P450 72A15-like	—	3,692	Downstream
TCONS_00069122	XLOC_037292	Cytochrome P450 72A13	—	33,818	Downstream
TCONS_00069410	Rs232150	JA-responsive protein 1	<i>JAC1</i>	7,495	Downstream
TCONS_00069424	Rs232150	JA-responsive protein 1	<i>JAC1</i>	70,884	Upstream
TCONS_00071696	XLOC_038932	Calcium-transporting ATPase 3, endoplasmic reticulum-type-like	—	−13,900	Antisense
TCONS_00079480	Rs306040	F-box family protein	<i>FBS1</i>	84,994	Upstream
TCONS_00096570	XLOC_053988	Receptor-like protein 12	<i>RLP12</i>	37,622	Downstream
TCONS_00110905	XLOC_073149	Probable LRR receptor-like serine/threonine-protein kinase	—	43,945	Downstream
TCONS_00123910	Rs478440	Auxin-responsive GH3 family protein	—	4,097	Downstream
TCONS_00126664	XLOC_079470	Transcription factor bHLH96-like	<i>bHLH96</i>	5,833	Downstream
	XLOC_083827	Autophagy-related protein 18h-like	—	41,217	Upstream
TCONS_00127122	XLOC_084067	Ethylene-responsive transcription factor ABR1-like	—	68,396	Upstream
TCONS_00138360	Rs510300	CBL-interacting protein kinase 25	<i>CIPK25</i>	3,660	Upstream
	XLOC_087207	Cytochrome P450 71B11	—	−4,846	Sense
TCONS_00149969	XLOC_095877	Enhanced disease resistance 2-like	—	513	Upstream

peptide processing), and antioxidant regulation (e.g., GO: 0000302: response to reactive oxygen species and GO: 0006979: response to oxidative stress). In the MF category, specific GO terms included numerous ion transport activities (e.g., GO: 0005216: ion channel activity; GO: 0005262: calcium channel activity; GO: 0005272: sodium channel activity; and GO: 0015079: potassium ion transmembrane transporter activity). Regarding the CC category, GO terms associated with the photosynthetic components (e.g., GO: 0009507: chloroplast; GO: 0009521: photosystem; GO: 0009522: photosystem I; and GO: 0009523: photosystem II) and ion transport channels (e.g., GO: 0034702: ion channel complex; GO: 0034703: cation channel

complex; and GO: 0034706: sodium channel complex) were enriched. Nevertheless, only one MF term “glutamate-cysteine ligase activity (GO: 0004357)” was significantly enriched in the NA_3_vs._CK comparison group for cis-acting targets.

The analysis conducted using KEGG suggested that the cis-acting target genes of DE-lncRNAs were highly represented in 119 KEGG pathways (Figure 3; Supplementary Table S9). Especially, the top 20 enriched pathways in each comparison group demonstrated significant enrichment of key pathways such as plant hormone signal transduction, homologous recombination, ubiquitin-mediated proteolysis, and zeatin biosynthesis.

TABLE 2 Representative differentially expressed trans-targets for some salt-responsive lncRNAs.

lncRNA_ID	mRNA_gene_ID	Description of the target gene	Gene name
TCONS_00007404	Rs069590	Lipoxygenase 3	<i>LOX3</i>
	Rs145870	Polygalacturonase inhibiting protein 1	<i>PGIP1</i>
	Rs159870	Trehalose-6-phosphate phosphatase	<i>TPPB</i>
	Rs232150	JA-responsive protein 1	<i>JAC1</i>
	Rs269180	TSK-associating protein 1	<i>TSA1</i>
	Rs375400	WRKY transcription factor 40	<i>WRKY40</i>
	Rs450390	Nitrate reductase 1	<i>NIA1</i>
	XLOC_072566	ABC transporter G family member 35	—
TCONS_00021574	Rs045300	Aminopeptidase M1	<i>APM1</i>
	Rs388430	Transcription factor MYB75	<i>MYB75</i>
TCONS_00024017	XLOC_021106	E3 ubiquitin-protein ligase RHA1B-like	—
TCONS_00031201	Rs038050	TSK-associating protein 1	<i>TSA1</i>
	Rs086530	Trehalose-6-phosphate phosphatase	<i>TPPB</i>
	Rs119710	Peroxidase	<i>POD</i>
	Rs228720	Leucine-rich repeat protein FLR1	<i>FLR1</i>
	XLOC_004367	ABC transporter G family member 35	—
TCONS_00036276	Rs251480	Water-soluble chlorophyll protein	<i>WSCP1</i>
	XLOC_076218	Abietadienol/abietadienal oxidase	—
TCONS_00036702	Rs142820	Protein TIFY 9	—
	Rs305010	Cytochrome P450	—
	Rs399730	Glutamate decarboxylase 1	<i>GAD1</i>
	XLOC_095772	Receptor-like protein kinase 2	<i>RPK2</i>
TCONS_00038561	Rs009970	ACC oxidase	—
	Rs290510	Vacuolar glucose transporter 1	<i>VGT1</i>
	Rs391910	Polygalacturonase inhibiting protein 1	<i>PGIP1</i>
	Rs424800	Transcription factor bHLH35	<i>bHLH35</i>
TCONS_00039084	XLOC_065872	Chalcone synthase 3	<i>CHS3</i>
TCONS_00043052	XLOC_001705	Zinc finger protein DOF1.3	<i>DOF1.3</i>
TCONS_00044336	XLOC_102899	Putative expansin-B2	<i>EXPB2</i>
TCONS_00047111	Rs096180	Pectate lyase	—
TCONS_00052720	Rs057150	Superoxide dismutase	<i>SOD</i>
TCONS_00056164	Rs155060	Xyloglucan endotransglucosylase/hydrolase 18	<i>XTH18</i>
	Rs306040	F-box family protein	<i>FBS1</i>
TCONS_00058325	XLOC_100431	Endoglucanase 19	—
TCONS_00060741	Rs236670	Transducin/WD40 repeat-like superfamily protein	—
	Rs478440	Auxin-responsive GH3 family protein	—
	Rs492760	Allene oxide cyclase 2	<i>AOC2</i>
TCONS_00071951	Rs392590	CBL-interacting protein kinase 2	<i>CIPK2</i>

(Continued on following page)

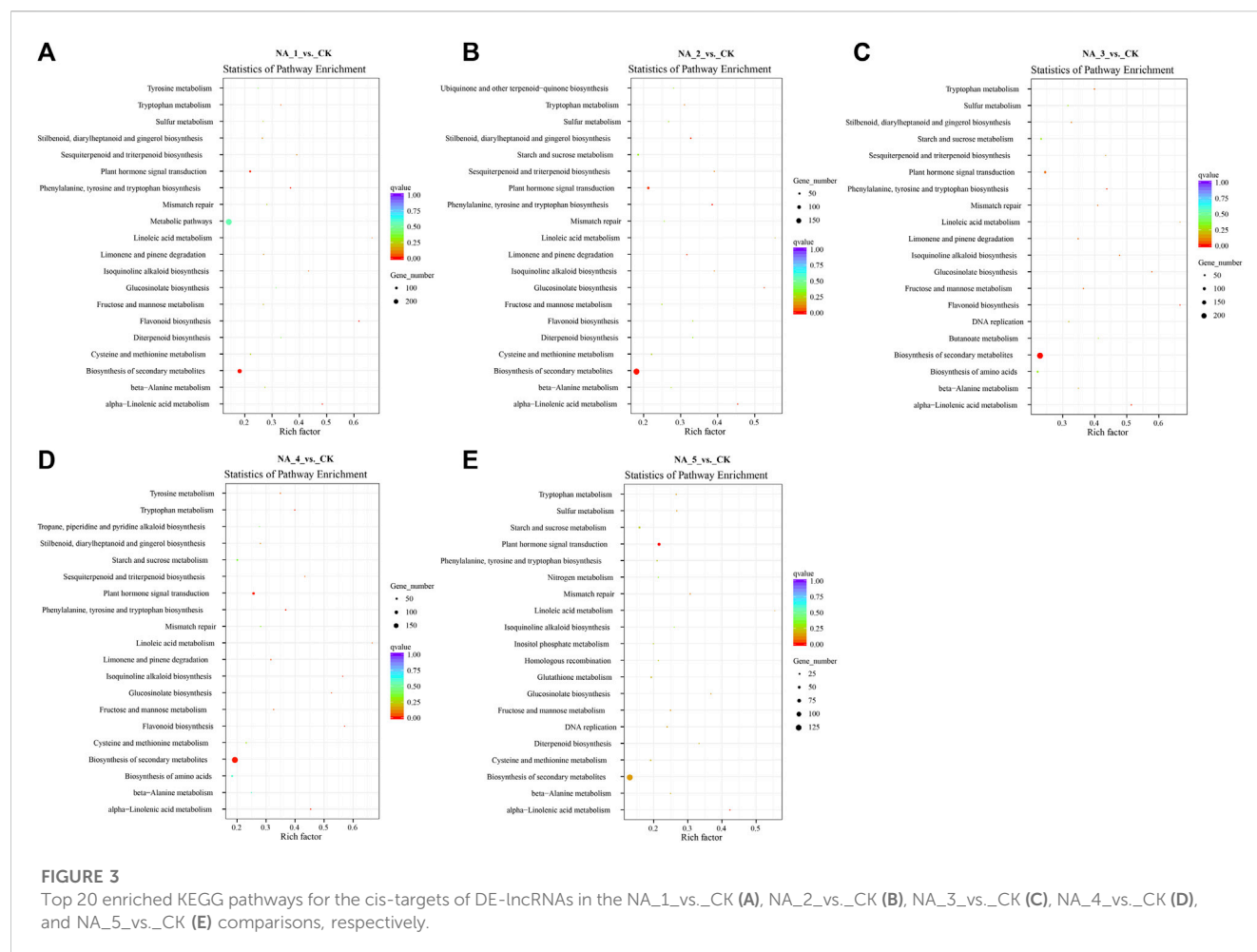
TABLE 2 (Continued) Representative differentially expressed trans-targets for some salt-responsive lncRNAs.

lncRNA_ID	mRNA_gene_ID	Description of the target gene	Gene name
	Rs412870	Lipoxygenase 3	<i>LOX3</i>
	Rs450300	Trehalose-6-phosphate phosphatase	<i>TPPB</i>
TCONS_00073919	Rs357900	Chitinase family protein	—
TCONS_00073924	XLOC_024931	MADS-box protein AGL24-like	—
TCONS_00081104	Rs211640	Cytochrome b5	—
TCONS_00094358	Rs074390	Pyruvate decarboxylase	<i>PDC1</i>
	Rs194210	bHLH transcription factor	—
TCONS_00100832	Rs492750	Allene oxide cyclase 3	<i>AOC3</i>
TCONS_00101865	XLOC_028048	Phytosulfokines 2	<i>PSK2_1</i>
	XLOC_041486	Calmodulin-like protein 11	<i>CML11</i>
TCONS_00115773	Rs119330	Annexin d4	<i>ANN4</i>
	Rs315450	Proline dehydrogenase	—
TCONS_00123532	XLOC_086136	Probable mannan synthase 7	—
TCONS_00133489	Rs069810	Lipoxygenase 3	<i>LOX3</i>
	Rs152410	Dehydration-responsive element-binding protein 3-like	—
	XLOC_027166	Squamosa promoter-binding-like protein 2	<i>SPL2</i>
TCONS_00137062	Rs006490	Sugar transporter EDR6-like 3	—
	Rs155070	Xyloglucan endotransglucosylase/hydrolase protein 20	<i>XTH20</i>
	Rs157300	Thioredoxin m4	—
TCONS_00138712	Rs132610	MATE efflux family protein	—
	Rs448590	WRKY transcription factor 40	<i>WRKY40</i>
TCONS_00140966	Rs043580	Mitogen-activated protein kinase kinase kinase 21	<i>MAPKKK21</i>
TCONS_00146159	Rs432680	Ethylene-responsive element-binding factor 1	<i>ERF1A</i>
TCONS_00147454	XLOC_012140	Chaperone protein dnaJ 8	—
TCONS_00152030	XLOC_075206	WD repeat domain-containing protein 83-like	—
TCONS_00152155	Rs501390	Proton-dependent oligopeptide transport family protein	—
TCONS_00155276	Rs269160	β -Glucosidase 18	<i>BGLU18</i>
TCONS_00159115	Rs067500	Sulfate transporter	—
	Rs159740	ACT domain repeat 7	<i>ACR7</i>
	Rs203420	Jacalin-related lectin 22	<i>JRL22</i>
	Rs458200	Phosphoinositide phospholipase C4	—
TCONS_00160055	Rs541640	Transducin/WD40 repeat-like superfamily protein	—
TCONS_00160242	XLOC_006145	CBL-interacting serine/threonine-protein kinase 17	<i>CIPK17</i>

Enrichment analysis for trans-acting target genes of DE-lncRNAs

Functional analysis suggested that the trans-acting target genes of DE-lncRNAs were enriched in 3,051 GO terms, including 1,710 BP terms, 418 MF terms, and 923 CC terms (Supplementary Table S10). As anticipated, the trans-targets of

DE-lncRNAs exhibited a significant overlap in GO terms pertaining to essential biological processes (e.g., photosynthesis, signal transduction, and antioxidant regulation), molecular functions (e.g., ion transport activity), and cellular components (e.g., photosynthetic component and ion transport channel) with the cis-acting targets. Furthermore, a notable enrichment of 11, 14, 18, 15, and 7 GO terms was



observed in the NA_1_vs._CK, NA_2_vs._CK, NA_3_vs._CK, NA_4_vs._CK, and NA_5_vs._CK comparisons, respectively (Supplementary Table S10). This demonstrates that the lncRNA-trans-regulatory target genes may play prior roles in response to salt stress in radish.

The trans-acting target genes of DE-lncRNAs exhibited enrichment in 118 KEGG pathways (the top 20 enriched pathways of each group are shown in Figure 4; Supplementary Table S11). Notably, several pathways, including α -linolenic acid metabolism, β -alanine metabolism, biosynthesis of secondary metabolites, fructose and mannose metabolism, glucosinolate biosynthesis, isoquinoline alkaloid biosynthesis, plant hormone signal transduction, and tryptophan metabolism, were significantly enriched across all comparison groups. Meanwhile, certain pathways associated with salt stress defense, such as plant hormone signal transduction, glutathione metabolism, ubiquinone, and other terpenoid-quinone biosynthesis, were also enriched for some target genes.

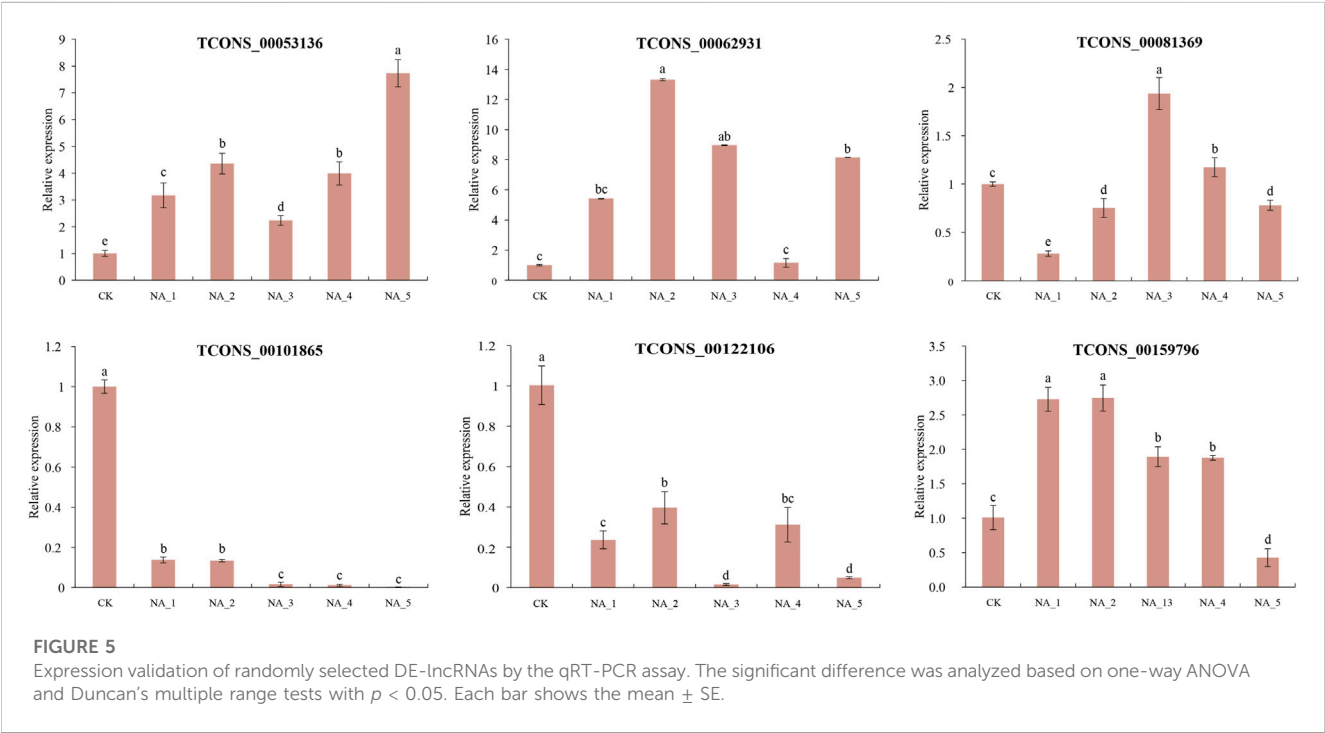
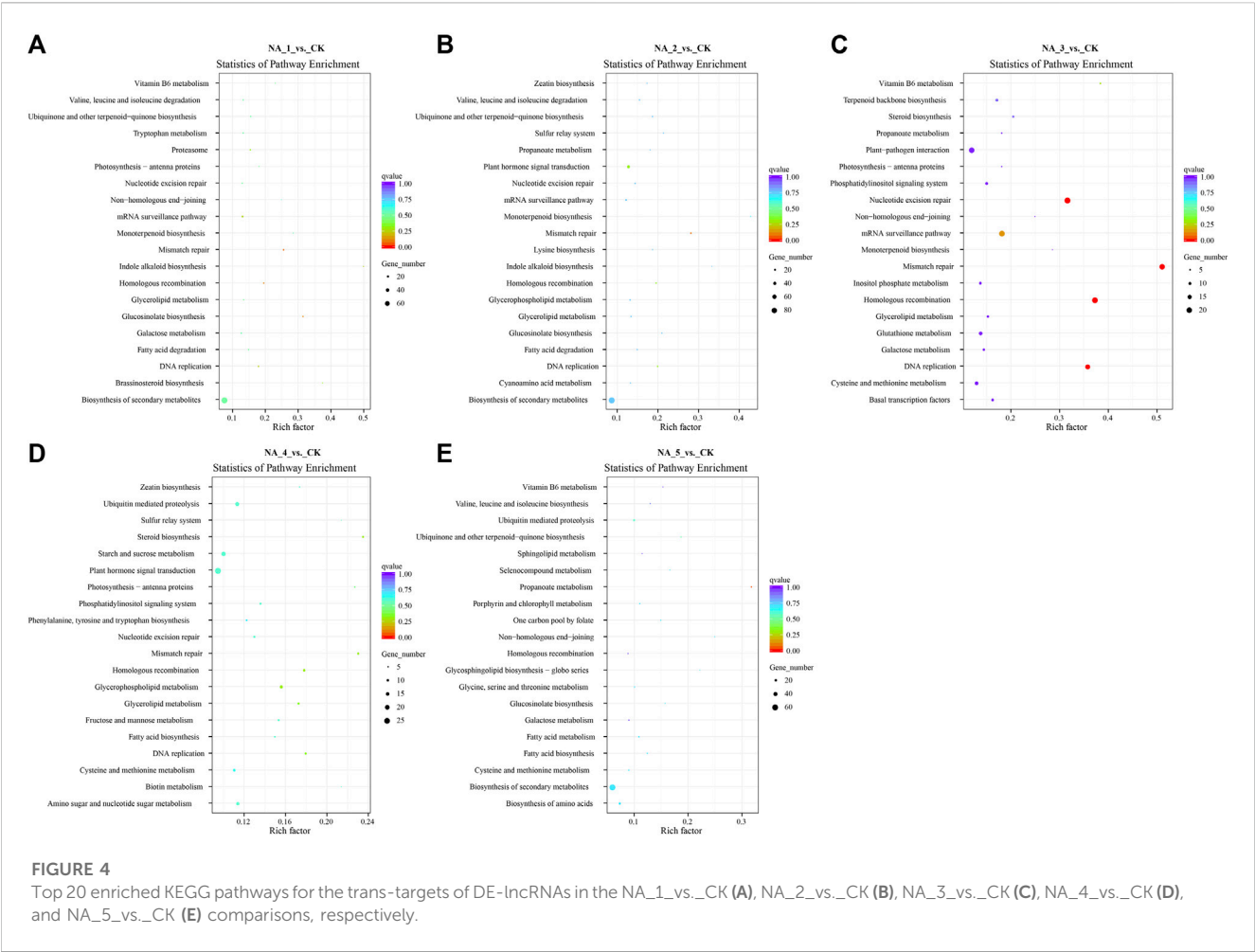
Experimental validation of DE-lncRNAs

To validate the analysis of the data obtained from RNA-seq, six lncRNA transcripts were examined by qRT-PCR (Figure 5). In general, the expression patterns of four salt-responsive lncRNAs

(TCONS_00053136, TCONS_00062931, TCONS_00101865, and TCONS_00122106) were relatively coincident with similar trends between qRT-PCR and RNA-seq, indicating the reliability of expression profiling based on RNA-seq data. However, the expression of TCONS_00081369 and TCONS_00159796 exhibited contrasting changes at certain time points during salt treatments when comparing the two methods. For instance, the upregulation of TCONS_00081369 at 12 h and 96 h of salt treatment was observed in the RNA-seq data, while a slight downregulation was detected in the qRT-PCR analysis. Conversely, TCONS_00159796 exhibited an opposite change in expression at 96 h under 200 mM NaCl. The observed disparity in expression levels between qRT-PCR and RNA-seq could be partially attributed to variations in sensitivity and algorithmic methodologies employed by these two approaches (Xu et al., 2015).

Discussion

Salt stress is considered to be one of the main constraints on radish yield and quality. Recently, numerous investigations have highlighted the primary correlation between regulatory function and salt stress-responsive lncRNAs in plants (Zhang et al., 2019; Fu et al., 2020; Liu et al., 2022a; Chen et al., 2023). However, the potential role of lncRNAs during salt stress conditions in radish remains



unexplored. In the current research, a total of 7,709 novel lncRNAs were systematically identified based on publicly accessible radish genome data using ssRNA-seq. Yet, in comparison to other plant species such as maize (Liu et al., 2022a), sesame (Gong et al., 2021), and peanut (Ma et al., 2020), only lincRNAs were observed in radish, manifesting that the lncRNA types may be highly specific to the context and species. Furthermore, it was also noticed that certain lncRNAs were not distributed on any chromosome, suggesting that these lncRNAs may be transcriptional noise (Nojima and Proudfoot, 2022). As anticipated, lncRNAs identified in radish were found to be shorter in total length and ORFs, and they also harbored fewer exons compared to protein-coding transcripts, which coincided with previous results in other plant species (Lv et al., 2016; Yan et al., 2020; Hu et al., 2022). We also recognized 363 lncRNAs that exhibit differential expression patterns in response to salt treatments (Supplementary Table S5). Notably, among these DE-lncRNAs, 33, 51, 72, 46, and 26 were exclusively detected at 6 h, 12 h, 24 h, 48 h, and 96 h of salt stress, respectively, when compared to the control (Figure 2A). These findings strongly suggested that the expression of salt-responsive lncRNAs was tightly regulated in a time-dependent manner, consistent with previous reports (Fu et al., 2020; Lee et al., 2022). A prior investigation has indicated that lncRNAs can exert their functions by enhancing gene expression at the transcriptional level (Ørom et al., 2010). Particularly, numerous upregulated DE-lncRNAs were dominant among all salinity conditions, implying a significant impact of high salinity on lncRNA expression.

Plants possess a diverse array of signaling molecules that facilitate the activation of appropriate defense responses in the presence of stress stimuli (Yu et al., 2017). The salt overly sensitive (SOS) signaling pathway, consisting of SOS3, SOS2, and SOS1 from the Na⁺/H⁺ exchanger (NHX), CBL-interacting protein kinase (CIPK), and calcineurin B-like (CBL) gene families, respectively, has been proposed to play a specific role in facilitating cellular signaling and maintaining ion homeostasis in response to salt stress in plants (Ji et al., 2013; Acharya et al., 2023). The CBL-CIPK signaling pathway, in particular, is a crucial component of the SOS pathway. In this work, although we did not directly predict SOS3, SOS2, and SOS1 as the lncRNA targets, the identification of salt-inducible DE-lncRNA-targeted CBL-interacting protein kinase 2 (CIPK2) and calcineurin B-like protein 1 (CBL1) might signify that lncRNAs play a role in regulating the SOS pathway in response to salt stress in radish. Annexin d4 (ANN4), a highly effective endogenous immunomodulatory protein, has been implicated in the cytosolic calcium signaling pathway in response to diverse stressors (Huh et al., 2010; Liao et al., 2017). In our experiment, the expression of TCONS_00115773-regulated ANN4 was significantly changed under salinity conditions (Table 2; Supplementary Table S7), indicating that lncRNA-ANN4-mediated calcium signaling may have a crucial function in enhancing radish resistance to high salinity. Phytohormones extensively participate in the adjustment of plants to unfavorable environmental elicitors. On the other hand, the intricate hormone signaling networks facilitate extensive crosstalk, thereby exerting crucial functions in the mediation of plant defense responses (Verma et al., 2016). Our analysis revealed the identification of DE-target genes involved in hormone signaling and responsive proteins, such as JA-responsive protein 1, and auxin-responsive GH3 family proteins (Tables 1, 2; Supplementary Table S7).

Furthermore, we also noted that DE-lncRNA targeted several genes encoding mitogen-activated protein kinase (MAPK) cascades, including MAPK20 and MAPKKK21 (Tables 1, 2; Supplementary Table S7). MAPK cascades are a set of vital signaling kinases that participate in the intracellular transmission of extracellular signals via fine-tuning specific TFs, functional proteins, and transporters (Zhang and Zhang, 2022). Importantly, it is suggested that the interplay between second messengers, hormones, and MAPK modules is instrumental in adapting to diverse environmental cues (Smékalová et al., 2014), suggesting the essential involvement of lncRNAs in sensing and transducing stress signals to activate the defense mechanisms of radish under salt stress conditions.

The interactions between stress-specific TFs and lncRNAs were hypothesized to play a significant role in response to various abiotic stresses, including salt stress (Mirdar Mansuri et al., 2022). For example, in wheat, Shumayla et al. (2017) demonstrated that lncRNAs were co-expressed with the TFs related to WRKY, NAC, MYB, ERF, C3H, C2H2, bZIP, and bHLH families under salt stress conditions; several lncRNAs associated with TFs belonging to ARF, C2C2(Zn), and HSF families were found in duckweed (Fu et al., 2020); in maize, 11 target transcripts of the salt-responsive lncRNAs belonging to seven TF families, including bHLH, C2H2, Hap3/NFYB, HAS, MYB, WD40, and WRKY, were predicted (Liu et al., 2022a). In this research, a specific group of genes related to TFs were also found in salt-responsive lncRNA-mRNA pairs in radish (Tables 1, 2; Supplementary Table S7). Specifically, the bHLH gene, *bHLH96*, was identified as a cis-regulated target gene of TCONS_00126664; a radish gene homologous to the *Arabidopsis* gene, namely, WRKY transcription factor 40 (*WRKY40*), was predicted to be targeted by TCONS_00007404 and TCONS_00138712. Moreover, TCONS_00146159 was identified to target a TF homologous to *Arabidopsis* *ERF1A*. The ERF gene *TdERF1* has been found to potentially play a role in the mechanism of salt susceptibility/tolerance by regulating multiple hormone signaling pathways in wheat (Makhloufi et al., 2014), indicating its potential involvement in modulating plant responses to salt constraints. In addition, a considerable number of genes from the TIFY transcription factor family were found to be targeted by DE-lncRNAs. A recent study has reported that TIFYs play an indispensable role in various aspects of plant biology, including growth, development, signal transduction, and response to stress stimuli in plants (Liu et al., 2022b).

A set of evidence also supported the fact that high salinity induces oxidative and osmotic limitations, leading to the accumulation of ROS and dehydration (Cabot et al., 2014; Yang and Guo, 2018; Xiao and Zhou, 2023). Our results indicated that certain targets of DE-lncRNAs related to antioxidants exhibited significant changes in expression levels when subjected to salt conditions (Table 2 and S6). For instance, TCONS_00052720 and TCONS_00031201 were found to fine-tune superoxide dismutase (*SOD*) and peroxidase (*POD*), respectively, which was consistent with the report in cotton (Zhang et al., 2019), suggesting the essential role of lncRNAs in scavenging excess ROS during salt stress. Salt stress also triggers reduced water availability for plants. In this study, several DE-lncRNAs were identified to target transcripts encoding proteins analogous to potato DREB1 with dehydration-responsive element-binding protein 3-like properties (Table 2; Supplementary Table S7), which had been

demonstrated to enhance salinity tolerance in transgenic potato plants (Bouaziz et al., 2013).

In addition, several functional proteins and transporters may play considerable roles in adjusting plant responses to salt stress. Chalcone synthase (CHS), a key enzyme in the flavonoid biosynthesis pathway, exerts a significant role in regulating plant growth, development, and defense against abiotic stress (Dao et al., 2011; Hou et al., 2022). The gene encoding chalcone synthase 3 (CHS3) was identified as a trans-acting target gene for four DE-lncRNAs (Table 2; Supplementary Table S7), suggesting that the lncRNA-mediated flavonoid biosynthesis pathway may play an essential role in response to salt stress in radish. Allene oxide cyclases (AOCs) were the pivotal genes in the JA biosynthetic pathway involved in regulating plant responses to developmental cues and environmental stresses. Numerous prior studies have demonstrated that the overexpression of allene oxide cyclase (AOC) enhances salinity tolerance in plants through the activation of jasmonate signaling (Pi et al., 2009; Zhao et al., 2014). In our study, a considerable number of DEGs encoding allene oxide cyclase 2 (AOC2) and allene oxide cyclase 3 (AOC3) were predicted to be potential targets of DE-lncRNAs (Table 2; Supplementary Table S7). Furthermore, the role of lipoxygenase 3 (LOX3) in *Arabidopsis* has been uncovered as a significant enzyme implicated in JA synthesis in response to salt stress (Ding et al., 2016). Our analysis differentiated that numerous DE-lncRNA-targeted genes encoding LOX3 displayed altered expression levels under salt stress conditions in radish (Table 2; Supplementary Table S7). Together, these findings might imply that the lncRNA-mediated regulation of the JA signaling pathway contributed to radish tolerance to salt stress in a complex and efficient way.

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.

References

- Acharya, B. R., Zhao, C., Reyes, L. A. R., Ferreira, J. F. S., and Sandhu, D. (2023). Understanding the salt overly sensitive pathway in *Prunus*: identification and characterization of NHX, CIPK, and CBL genes. *Plant Genome*, e20371. doi:10.1002/tpg2.20371
- Baruah, P. M., Kashyap, P., Krishnatreya, D. B., Bordoloi, K. S., and Agarwala, N. (2021). Identification and functional analysis of drought responsive lncRNAs in tea plant. *Plant gene*, 27, 100311. doi:10.1016/j.plgene.2021.100311
- Bouaziz, D., Pirello, J., Charfeddine, M., Hammami, A., Jbir, R., Dhieb, A., et al. (2013). Overexpression of StDREB1 transcription factor increases tolerance to salt in transgenic potato plants. *Mol. Biotechnol.* 54, 803–817. doi:10.1007/s12033-012-9628-2
- Cabot, C., Sibole, J. V., Barceló, J., and Poschenrieder, C. (2014). Lessons from crop plants struggling with salinity. *Plant Sci.* 226, 2–13. doi:10.1016/j.plantsci.2014.04.013
- Cao, W., Gan, L., Wang, C., Zhao, X., Zhang, M., Du, J., et al. (2021). Genome-wide identification and characterization of potato long non-coding RNAs associated with *Phytophthora infestans* resistance. *Front. Plant Sci.* 12, 619062. doi:10.3389/fpls.2021.619062
- Chekanova, J. A. (2015). Long non-coding RNAs and their functions in plants. *Curr. Opin. Plant Biol.* 27, 207–216. doi:10.1016/j.pbi.2015.08.003
- Chen, X., Jiang, X., Niu, F., Sun, X., Hu, Z., Gao, F., et al. (2023). Overexpression of lncRNA77580 regulates drought and salinity stress responses in soybean. *Plants (Basel)*. 12, 181. doi:10.3390/plants12010181
- Chen, X., Sun, Y., Cai, R., Wang, G., Shu, X., and Pang, W. (2018). Long noncoding RNA: multiple players in gene expression. *BMB Rep.* 51, 280–289. doi:10.5483/bmbrep.2018.51.6.025
- Dao, T. T., Linthorst, H. J., and Verpoorte, R. (2011). Chalcone synthase and its functions in plant resistance. *Phytochem. Rev.* 10, 397–412. doi:10.1007/s11101-011-9211-7
- Ding, H., Lai, J., Wu, Q., Zhang, S., Chen, L., Dai, Y. S., et al. (2016). Jasmonate complements the function of *Arabidopsis* lipoxygenase 3 in salinity stress response. *Plant Sci.* 244, 1–7. doi:10.1016/j.plantsci.2015.11.009
- Fu, L., Ding, Z., Tan, D., Han, B., Sun, X., and Zhang, J. (2020). Genome-wide discovery and functional prediction of salt-responsive lncRNAs in duckweed. *BMC Genomics* 21, 212. doi:10.1186/s12864-020-6633-x
- Gil, N., and Ulitsky, I. (2020). Regulation of gene expression by cis-acting long non-coding RNAs. *Nat. Rev. Genet.* 21, 102–117. doi:10.1038/s41576-019-0184-5
- Gong, H., You, J., Zhang, X., Liu, Y., Zhao, F., Cui, X., et al. (2021). Genome-wide identification and functional analysis of long non-coding RNAs in sesame response to salt stress. *J. Plant Biol.* 64, 555–565. doi:10.1007/s12374-021-09324-3
- Grattan, S. R. (2016). *Drought tip: crop salt tolerance*. Richmond, CA, USA: University of California, Agriculture and Natural Resources.
- Hou, Q., Li, S., Shang, C., Wen, Z., Cai, X., Hong, Y., et al. (2022). Genome-wide characterization of chalcone synthase genes in sweet cherry and functional characterization of CpCHS1 under drought stress. *Front. Plant Sci.* 13, 989959. doi:10.3389/fpls.2022.989959
- Hou, X., Cui, J., Liu, W., Jiang, N., Zhou, X., Qi, H., et al. (2020). LncRNA39026 enhances tomato resistance to *Phytophthora infestans* by decoying miR168a and inducing PR gene expression. *Phytopathology* 110, 873–880. doi:10.1094/PHYTO-12-19-0445-R

Author contributions

XS conceived and designed the research. MT and YS performed the experiments. WD and GC contributed powerful analytical tools. XS, LX, and XL analyzed data. XS wrote the manuscript. MT, LX, ZH, and CJ revised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was, in part, supported by grants from the National Natural Science Foundation of China (32102399) and the Natural Science Foundation of Jiangsu Province (BK20181062).

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/fgene.2023.1232363/full#supplementary-material>

- Hu, X., Wei, Q., Wu, H., Huang, Y., Peng, X., Han, G., et al. (2022). Identification and characterization of heat-responsive lncRNAs in maize inbred line CM1. *BMC Genomics* 23, 208. doi:10.1186/s12864-022-08448-1
- Huang, X., Zhang, H., Wang, Q., Guo, R., Wei, L., Song, H., et al. (2021). Genome-wide identification and characterization of long non-coding RNAs involved in flag leaf senescence of rice. *Plant Mol. Biol.* 105, 655–684. doi:10.1007/s11103-021-01121-3
- Huh, S. M., Noh, E. K., Kim, H. G., Jeon, B. W., Bae, K., Hu, H. C., et al. (2010). Arabidopsis annexins AnnAt1 and AnnAt4 interact with each other and regulate drought and salt stress responses. *Plant Cell Physiol.* 51, 1499–1514. doi:10.1093/pcp/pcq111
- Islam, W., Waheed, A., Naveed, H., and Zeng, F. (2022). MicroRNAs mediated plant responses to salt stress. *Cells* 11, 2806. doi:10.3390/cells11182806
- Jha, U. C., Nayyar, H., Jha, R., Khurshid, M., Zhou, M., Mantri, N., et al. (2020). Long non-coding RNAs: emerging players regulating plant abiotic stress response and adaptation. *BMC Plant Biol.* 20, 466. doi:10.1186/s12870-020-02595-x
- Ji, H., Pardo, J. M., Batelli, G., Van Oosten, M. J., Bressan, R. A., and Li, X. (2013). The Salt Overly Sensitive (SOS) pathway: established and emerging roles. *Mol. Plant* 6, 275–286. doi:10.1093/mp/sst017
- Kapranov, P., Cheng, J., Dike, S., Nix, D. A., Duttagupta, R., Willingham, A. T., et al. (2007). RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science* 316, 1484–1488. doi:10.1126/science.1138341
- Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 12, 357–360. doi:10.1038/nmeth.3317
- Lee, M. H., Kim, K. M., Sang, W. G., Kang, C. S., and Choi, C. (2022). Comparison of gene expression changes in three wheat varieties with different susceptibilities to heat stress using RNA-Seq analysis. *Int. J. Mol. Sci.* 23, 10734. doi:10.3390/ijms231810734
- Liao, C., Zheng, Y., and Guo, Y. (2017). MYB30 transcription factor regulates oxidative and heat stress responses through ANNEXIN-mediated cytosolic calcium signaling in Arabidopsis. *New Phytol.* 216, 163–177. doi:10.1111/nph.14679
- Liu, P., Zhang, Y., Zou, C., Yang, C., Pan, G., Ma, L., et al. (2022a). Integrated analysis of long non-coding RNAs and mRNAs reveals the regulatory network of maize seedling root responding to salt stress. *BMC Genomics* 23, 50. doi:10.1186/s12864-021-08286-7
- Liu, X., Yu, F., Yang, G., Liu, X., and Peng, S. (2022b). Identification of TIFY gene family in walnut and analysis of its expression under abiotic stresses. *BMC Genomics* 23, 190. doi:10.1186/s12864-022-08416-9
- 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
- Lv, Y., Liang, Z., Ge, M., Qi, W., Zhang, T., Lin, F., et al. (2016). Genome-wide identification and functional prediction of nitrogen-responsive intergenic and intronic long non-coding RNAs in maize (*Zea mays* L.). *BMC Genomics* 17, 350. doi:10.1186/s12864-016-2650-1
- Ma, L., Bajic, V. B., and Zhang, Z. (2013). On the classification of long non-coding RNAs. *RNA Biol.* 10, 925–933. doi:10.4161/rna.24604
- Ma, X., Zhang, X., Traore, S. M., Xin, Z., Ning, L., Li, K., et al. (2020). Genome-wide identification and analysis of long noncoding RNAs (lncRNAs) during seed development in peanut (*Arachis hypogaea* L.). *BMC Plant Biol.* 20, 192. doi:10.1186/s12870-020-02405-4
- Makhoulfi, E., Yousfi, F. E., Marande, W., Mila, I., Hanana, M., Bergès, H., et al. (2014). Isolation and molecular characterization of ERF1, an ethylene response factor gene from durum wheat (*Triticum turgidum* L. subsp. durum), potentially involved in salt-stress responses. *J. Exp. Bot.* 65, 6359–6371. doi:10.1093/jxb/eru352
- Mirdar Mansuri, R., Azizi, A. H., Sadri, A. H., and Shobbar, Z. S. (2022). Long non-coding RNAs as the regulatory hubs in rice response to salt stress. *Sci. Rep.* 12, 21696. doi:10.1038/s41598-022-26133-x
- Nojima, T., and Proudfoot, N. J. (2022). Mechanisms of lncRNA biogenesis as revealed by nascent transcriptomics. *Nat. Rev. Mol. Cell Biol.* 23, 389–406. doi:10.1038/s41580-021-00447-6
- Ørom, U. A., Derrien, T., Guigo, R., and Shiekhattar, R. (2010). Long noncoding RNAs as enhancers of gene expression. *Quant. Biol.* 75, 325–331. doi:10.1101/sqb.2010.75.058
- 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
- Pi, Y., Jiang, K., Cao, Y., Wang, Q., Huang, Z., Li, L., et al. (2009). Allene oxide cyclase from *Camptotheca acuminata* improves tolerance against low temperature and salt stress in tobacco and bacteria. *Mol. Biotechnol.* 41, 115–122. doi:10.1007/s12033-008-9106-z
- Quan, M., Chen, J., and Zhang, D. (2015). Exploring the secrets of long noncoding RNAs. *Int. J. Mol. Sci.* 16, 5467–5496. doi:10.3390/ijms16035467
- Safdar, H., Amin, A., Shafiq, Y., Ali, A., and Yasin, R. (2019). A review: impact of salinity on plant growth. *Nat. Sci.* 17, 34–40. doi:10.7537/marsnsj170119.06
- Shin, W. J., Nam, A. H., Kim, J. Y., Kwak, J. S., Song, J. T., and Seo, H. S. (2022). Intronic long noncoding RNA, RICE FLOWERING ASSOCIATED (RIFLA), regulates OsMADS56-mediated flowering in rice. *Plant Sci.* 320, 111278. doi:10.1016/j.plantsci.2022.111278
- Shumayla, S., Sharma, S., Taneja, M., Tyagi, S., Singh, K., and Upadhyay, S. K. (2017). Survey of high throughput RNA-Seq data reveals potential roles for lncRNAs during development and stress response in bread wheat. *Front. Plant Sci.* 8, 1019. doi:10.3389/fpls.2017.01019
- Směkalová, V., Doskočilová, A., Komis, G., and Samaj, J. (2014). Crosstalk between secondary messengers, hormones and MAPK modules during abiotic stress signalling in plants. *Biotechnol. Adv.* 32, 2–11. doi:10.1016/j.biotechadv.2013.07.009
- St Laurent, G., Wahlestedt, C., and Kapranov, P. (2015). The Landscape of long noncoding RNA classification. *Trends Genet.* 31, 239–251. doi:10.1016/j.tig.2015.03.007
- Sun, X., Xu, L., Wang, Y., Luo, X., Zhu, X., Kinuthia, K. B., et al. (2016). Transcriptome-based gene expression profiling identifies differentially expressed genes critical for salt stress response in radish (*Raphanus sativus* L.). *Plant Cell Rep.* 35, 329–346. doi:10.1007/s00299-015-1887-5
- Sun, X., Xu, L., Wang, Y., Yu, R., Zhu, X., Luo, X., et al. (2015). Identification of novel and salt-responsive miRNAs to explore miRNA-mediated regulatory network of salt stress response in radish (*Raphanus sativus* L.). *BMC Genomics* 16, 197. doi:10.1186/s12864-015-1416-5
- Sun, Z., Huang, K., Han, Z., Wang, P., and Fang, Y. (2020). Genome-wide identification of Arabidopsis long noncoding RNAs in response to the blue light. *Sci. Rep.* 10, 6229. doi:10.1038/s41598-020-63187-1
- Tian, Y., Bai, S., Dang, Z., Hao, J., Zhang, J., and Hasi, A. (2019). Genome-wide identification and characterization of long non-coding RNAs involved in fruit ripening and the climacteric in Cucumis melo. *BMC Plant Biol.* 19, 369. doi:10.1186/s12870-019-1942-4
- Verma, V., Ravindran, P., and Kumar, P. P. (2016). Plant hormone-mediated regulation of stress responses. *BMC Plant Biol.* 16, 86. doi:10.1186/s12870-016-0771-y
- Wang, H., Chung, P. J., Liu, J., Jang, I. C., Kean, M. J., Xu, J., et al. (2014). Genome-wide identification of long noncoding natural antisense transcripts and their responses to light in Arabidopsis. *Genome Res.* 24, 444–453. doi:10.1101/gr.165555.113
- Wang, H. V., and Chekanova, J. A. (2017). Long noncoding RNAs in plants. *Adv. Exp. Med. Biol.* 1008, 133–154. doi:10.1007/978-981-10-5203-3_5
- Wang, K. C., and Chang, H. Y. (2011). Molecular mechanisms of long noncoding RNAs. *Mol. Cell* 43, 904–914. doi:10.1016/j.molcel.2011.08.018
- Wierzbicki, A. T., Blevins, T., and Swiezewski, S. (2021). Long noncoding RNAs in plants. *Annu. Rev. Plant Biol.* 72, 245–271. doi:10.1146/annurev-arplant-093020-035446
- Xiao, F., and Zhou, H. (2023). Plant salt response: perception, signaling, and tolerance. *Front. Plant Sci.* 13, 1053699. doi:10.3389/fpls.2022.1053699
- Xu, L., Wang, Y., Liu, W., Wang, J., Zhu, X., Zhang, K., et al. (2015). *De novo* sequencing of root transcriptome reveals complex cadmium-responsive regulatory networks in radish (*Raphanus sativus* L.). *Plant Sci.* 236, 313–323. doi:10.1016/j.plantsci.2015.04.015
- Yan, X., Ma, L., and Yang, M. (2020). Identification and characterization of long non-coding RNA (lncRNA) in the developing seeds of *Jatropha curcas*. *Sci. Rep.* 10, 10395. doi:10.1038/s41598-020-67410-x
- Yang, Y., and Guo, Y. (2018). Elucidating the molecular mechanisms mediating plant salt-stress responses. *New Phytol.* 217, 523–539. doi:10.1111/nph.14920
- Yu, X., Feng, B., He, P., and Shan, L. (2017). From chaos to harmony: responses and signaling upon microbial pattern recognition. *Annu. Rev. Phytopathol.* 55, 109–137. doi:10.1146/annurev-phyto-080516-035649
- Zhang, M., and Zhang, S. (2022). Mitogen-activated protein kinase cascades in plant signaling. *J. Integr. Plant Biol.* 64, 301–341. doi:10.1111/jipb.13215
- Zhang, X., Dong, J., Deng, F., Wang, W., Cheng, Y., Song, L., et al. (2019). The long non-coding RNA lncRNA973 is involved in cotton response to salt stress. *BMC Plant Biol.* 19, 459. doi:10.1186/s12870-019-2088-0
- Zhang, Y., Fan, F., Zhang, Q., Luo, Y., Liu, Q., Gao, J., et al. (2022). Identification and functional analysis of long non-coding RNA (lncRNA) in response to seed aging in rice. *Plants (Basel)* 11, 3223. doi:10.3390/plants11233223
- Zhang, Y. C., Liao, J. Y., Li, Z. Y., Yu, Y., Zhang, J. P., Li, Q. F., et al. (2014). Genome-wide screening and functional analysis identify a large number of long noncoding RNAs involved in the sexual reproduction of rice. *Genome Biol.* 15, 512. doi:10.1186/s13059-014-0512-1
- Zhao, Y., Dong, W., Zhang, N., Ai, X., Wang, M., Huang, Z., et al. (2014). A wheat allene oxide cyclase gene enhances salinity tolerance via jasmonate signaling. *Plant Physiol.* 164, 1068–1076. doi:10.1104/pp.113.227595
- Zhou, D., Chen, C., Jin, Z., Chen, J., Lin, S., Lyu, T., et al. (2022a). Transcript profiling analysis and ncRNAs' identification of male-sterile systems of Brassica campestris reveal new insights into the mechanism underlying anther and pollen development. *Front. Plant Sci.* 13, 806865. doi:10.3389/fpls.2022.806865
- Zhou, H., Ren, F., Wang, X., Qiu, K., Sheng, Y., Xie, Q., et al. (2022b). Genome-wide identification and characterization of long noncoding RNAs during peach (*Prunus persica*) fruit development and ripening. *Sci. Rep.* 12, 11044. doi:10.1038/s41598-022-15330-3



OPEN ACCESS

EDITED BY

Krishnanand P. Kulkarni,
Delaware State University, United States

REVIEWED BY

Qin Yang,
Northwest A&F University, China
Zhenhua Wang,
Northeast Agricultural University, China

*CORRESPONDENCE

Mathews M. Dida,
✉ mitodida@yahoo.com
Manje Gowda,
✉ m.gowda@cgiar.org

RECEIVED 24 August 2023

ACCEPTED 18 October 2023

PUBLISHED 07 November 2023

CITATION

Omondi DO, Dida MM, Berger DK,
Beyene Y, Nsibo DL, Juma C,
Mahabaleswara SL and Gowda M (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. 14:1282673.
doi: 10.3389/fgene.2023.1282673

COPYRIGHT

© 2023 Omondi, Dida, Berger, Beyene,
Nsibo, Juma, Mahabaleswara and Gowda.
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.

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

Dennis O. Omondi^{1,2}, Mathews M. Dida^{1*}, Dave K. Berger³,
Yoseph Beyene⁴, David L. Nsibo³, Collins Juma^{4,2},
Suresh L. Mahabaleswara⁴ and Manje Gowda^{4*}

¹Department of Crops and Soil Sciences, School of Agriculture, Food Security and Environmental Sciences, Maseno University, Kisumu, Kenya, ²Crop Science Division Bayer East Africa Limited, Nairobi, Kenya, ³Department of Plant and Soil Sciences, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Pretoria, South Africa, ⁴The Global Maize Program, International Maize and Wheat Improvement Center (CIMMYT), Nairobi, Kenya

Among the diseases threatening maize production in Africa are gray leaf spot (GLS) caused by *Cercospora zeina* and northern corn leaf blight (NCLB) caused by *Exserohilum turcicum*. The two pathogens, which have high genetic diversity, reduce the photosynthesizing ability of susceptible genotypes and, hence, reduce the grain yield. To identify population-based quantitative trait loci (QTLs) for GLS and NCLB resistance, a biparental population of 230 lines derived from the tropical maize parents CML511 and CML546 and an association mapping panel of 239 tropical and sub-tropical inbred lines were phenotyped across multi-environments in western Kenya. Based on 1,264 high-quality polymorphic single-nucleotide polymorphisms (SNPs) in the biparental population, we identified 10 and 18 QTLs, which explained 64.2% and 64.9% of the total phenotypic variance for GLS and NCLB resistance, respectively. A major QTL for GLS, *qGLS1_186* accounted for 15.2% of the phenotypic variance, while *qNCLB3_50* explained the most phenotypic variance at 8.8% for NCLB resistance. Association mapping with 230,743 markers revealed 11 and 16 SNPs significantly associated with GLS and NCLB resistance, respectively. Several of the SNPs detected in the association panel were co-localized with QTLs identified in the biparental population, suggesting some consistent genomic regions across genetic backgrounds. These would be more relevant to use in field breeding to improve resistance to both diseases. Genomic prediction models trained on the biparental population data yielded average prediction accuracies of 0.66–0.75 for the disease traits when validated in the same population. Applying these prediction models to the association panel produced accuracies of 0.49 and 0.75 for GLS and NCLB, respectively. This research conducted in maize fields relevant to farmers in western Kenya has combined linkage and association mapping to identify new QTLs and confirm previous QTLs for GLS and NCLB resistance. Overall, our

findings imply that genetic gain can be improved in maize breeding for resistance to multiple diseases including GLS and NCLB by using genomic selection.

KEYWORDS

maize, gray leaf spot, northern corn leaf blight, quantitative trait loci, association mapping, genome-wide association study

1 Introduction

Despite its importance, maize production in Kenya is still low with an estimated average production of 1.8 t/ha⁻¹ among smallholder farmers when compared to the country's potential production yield of 6 t/ha⁻¹ (Munialo et al., 2019). This is partly due to the threat of highly destructive and virulent fungal pathogens limiting crop production (Beyene et al., 2019). In this context, northern corn leaf blight (NCLB), also known as northern leaf blight (NLB) or *Turcicum* leaf blight (TLB), caused by *Exserohilum turcicum* (Pass.) (Leonard and Suggs, 1974), and gray leaf spot (GLS) caused by *Cercospora zeina* Crous & U Braun (Crous et al., 2006) on the African continent (Nsibo et al., 2019; Nsibo et al., 2021) are the most lethal and economically significant foliar diseases of maize (Beyene et al., 2019; Sserumaga et al., 2020).

The two diseases reduce the photosynthetic potential of a plant and eventually decrease the grain yield (Saito et al., 2018). GLS is characterized by tan-to-gray rectangular lesions that are limited within the leaf veins (Korsman et al., 2012). It is associated with yield losses of approximately 30%–50%, particularly when using susceptible lines (Kinyua et al., 2010). On the other hand, NCLB is characterized by long elliptical cigar-shaped lesions on leaves that are gray-green (Welz, 1998). NCLB or TLB has been reported to cause yield reductions of 36%–72% in susceptible maize genotypes (Berger et al., 2020). The high genetic diversity reported for *C. zeina* and *E. turcicum* in Kenya (Borchardt et al., 1998; Nsibo et al., 2021) could lead to recombining pathogen populations, hence posing a greater risk to the vulnerable susceptible lines (McDonald and Linde, 2002). Therefore, there is a need to continuously discover new sources of resistance.

The doubled haploid (DH) technology offers the fastest alternative to achieve 100% homozygosity (attained in two generations) which is essential for a mapping population and population improvement (Prasanna et al., 2021). To complement the DH technology, genotyping by sequencing platforms such as Diversity Arrays Technology (DArT) offers a high-throughput platform for genotyping single-nucleotide polymorphism (SNP) markers (Kilian et al., 2012; Sansaloni et al., 2020). The DArTseq platform is purposefully a powerful tool for genome-wide discovery of SNP markers without prior sequence information (Wenzl et al., 2004). In addition, it generates high-density linkage maps, and it is also cost-competitive (Jaccoud et al., 2001; Sánchez-Sevilla et al., 2015).

Complex traits such as GLS and NCLB resistance are controlled by polygenic genes with minor effects that are distributed throughout the genome (Welz and Geiger, 2000; Wissner et al., 2006; Poland et al., 2011; Van Inghelandt et al., 2012; Chen et al., 2015; Ding et al., 2015). Mapping of the quantitative trait loci (QTLs) based on linkage analysis is a powerful tool for identifying the genomic regions associated

with the traits of interest. Previous QTL studies have mapped QTL for resistance to GLS and NCLB on all 10 maize chromosomes (Berger et al., 2014; Chen et al., 2015). A number of these QTLs have been fine-mapped with others cloned and the molecular mechanisms underlying such QTL characterized. Additionally, QTL mapping offers the advantage of mapping as early as at the F₂ populations; however, this is characterized by the limited number of recombination events captured and sizeable confidence interval (Challa and Neelapu, 2018; Rashid et al., 2020).

Genome-wide association studies (GWAS) attempt to overcome the drawbacks of QTL mapping as they utilize age-old recombination events in a large array of unrelated individuals leading to high-speed decay of linkage disequilibrium (Xiao et al., 2017; Kolkman et al., 2020). GWAS studies dig into the entire genome of different varieties (considering the SNPs present in the genotypic data) to establish the link between genotypic variations and the corresponding trait (Challa and Neelapu, 2018). Kibe et al. (2020a) combined the use of linkage mapping and GWAS to detect the significant SNPs and QTL conditioning resistance to GLS in an Improved Maize for African Soils (IMAS) diversity panel and a set of DH populations in Kenya. Several putative candidate genes involved in the transportation channel were identified to have a role in plant defense. In the present study, we attempted to validate some of the GLS resistance QTLs reported in the study by Kibe et al. (2020a) by using a common tropical parent CML511.

Genomic prediction (GP) is another promising genomic tool that has been applied successfully in plant breeding programs (Crossa et al., 2017). Previous studies indicated the potential of GP to increase genetic gain and reduce the time taken in breeding programs significantly (Beyene et al., 2019; Kibe et al., 2020a; Kibe et al., 2020b). In contrast to genetic mapping which identifies significant marker-trait associations, GP uses all markers available to estimate their effects, thus providing a powerful approach to account for any effects that might have been missed by either genetic or association mapping (Beyene et al., 2019). GP exhibits a close relationship to GWAS owing to the large genomic and phenotypic datasets used by the methods (Beyene et al., 2021).

However, this does not mean the complete withdrawal of genetic mapping but rather the incorporation of the two in genetic studies as complementary approaches since each provides considerable advantages. With this background, the objectives of this study were as follows: (1) to phenotypically characterize an elite tropical DH population and 240 tropical and sub-tropical maize inbred lines panel for their responses to GLS and NCLB, including correlation with other agronomic traits; (2) to identify population-based common QTL regions and significant SNPs using GWAS and linkage mapping; and (3) to assess the usefulness of GP in breeding for GLS and NCLB resistance in tropical maize.

2 Materials and methods

2.1 Study sites and genetic material

This study used (i) a biparental DH population derived from the tropical × tropical germplasm CML511 × CML546 inbred lines and (ii) an association panel made up of a collection of 239 tropical and sub-tropical maize inbred lines with early and intermediate maturity in Eastern Africa, representing some of the genetic diversity (for low N, drought, and biotic stresses, [Beyene et al., 2021](#); [Prasanna et al., 2021](#)). The association panel was evaluated in three locations in western Kenya, at Kitale (1.0191° N and 35.0023° E, 1900 masl); Shikutsa (0°16'57.83"N and 34°45'6.71"E, 1561 masl); and Kakamega (0°17'3.19"N and 34°45'8.24"E, 1535 masl). The biparental population was evaluated across different ecologies in western Kenya; Maseno University field demonstration site in 2018 (0°00'18.2"S and 34°35'43.5"E, 1500 masl), Maseno 2019 (0°00'07.0"S and 34°35'41.9"E, 1503 masl), and farmer's field in Kabianga 2018 (0°25'24.1"S and 35°07'31.7"E, 1780 masl).

2.2 Experimental design

Two hundred and thirty (230) entries (228 DH lines and two parents) of the biparental population were planted in a 5 × 46 alpha lattice design, randomized, and replicated three times at each site by using the CIMMYT's field book ([Vivek et al., 2007](#)). The association panel was also planted in 5 × 48 alpha lattice design, randomized, and replicated two times in each of three environments. Experimental plots consisted of 3 m long single rows with the rows spaced at 0.6 m apart. Adjacent plots were planted 0.75 m apart with an alley of 1.2 m at the end of each plot. Each plot was planted with 13 hills, with two seeds getting planted per hill. Thinning was later conducted to one plant per hill. Border rows of susceptible genotypes were also planted to act as spreaders of the pathogen. The experimental plots were managed using standard agronomic practices.

2.3 Phenotypic evaluation and data collection

GLS and NCLB disease severity (DS) were scored on a per-row basis using an ordinal 1–9 adapted from the work of [Berger et al. \(2014\)](#) for GLS and [Berger et al. \(2020\)](#) for NCLB. For DH population, DS ratings for GLS and NCLB were taken once per week for at least 5 weeks starting on average at 15 days after flowering (R2; reproductive stage two). All the data were collected using the CIMMYT's field book ([Vivek et al., 2007](#)). The DS scale was used as follows: score 1 for no GLS or NCLB lesions visible on the entire plant, 2 indicated close inspection of each leaf is necessary to find lesions, 3 indicated lesions are more easily seen but are majorly restricted to leaves lying below the ears, 4 indicated individual lesions are just becoming visible on the ear leaf and the leaves above the ears, 5 indicated lesions are more visible on the leaves above the ears, with the infections capturing <10% of the top leaves, 6 indicated lesions are more easily seen on the leaves above the ear leaf with infections

covering >10% of the leaf area, 7 for GLS and NCLB lesions dominating the leaf area on all the leaves with 50% of the maize leaf surface diseased, 8 for GLS and NCLB lesions prevalent on all the leaves of the maize plant with 80% of the maize leaf surface diseased, and 9 for GLS and NCLB lesions prevalent on all the leaves of the maize plant with the maize plant exhibiting a gray appearance with >80% of the maize leaf area diseased. For both GLS and NCLB, the DS scores over five intervals were used to calculate the area under the disease progress curve (AUDPC, [Shaner, 1977](#)). For the GWAS panel, both DS data were recorded at the dough stage of the plants. For DH population, data were also collected on days to anthesis (AD, the number of days from planting to when 50% of the plants in a plot were shedding pollen), days to silking (SD, the number of days from planting to when 50% of maize crops in a plot were showing silk), plant height (PH, cm), and ear height (EH, cm). Maize development stages were recorded using the scale of Purdue University (<http://extension.entm.purdue.edu/fieldcropsipm/corn-stages.php>).

2.4 Statistical analysis of the phenotypic data

Multi-environment trial analysis with R for windows (META-R) version 6.0 ([Alvarado et al., 2015](#)) was used to obtain the best linear unbiased estimations (BLUEs) and best linear unbiased predictions (BLUPs). In addition to BLUEs and BLUPs, META-R was also used to compute the genetic correlations among all the variables and among environments, least significant difference (LSD), grand mean, variance components, coefficients of variation (CV), and broad-sense heritability for all the variables. Analysis of the phenotypic data for both biparental population and association panel was conducted both within and across environments.

The BLUEs and the BLUPs were calculated for DS of GLS and NCLB, PH, EH, AD, SD, and the AUDPC which were the response variables. The columns in the input files were selected to be the factor names with the environment, replicate, block, and genotype as the independent variables. For analysis across environments using a lattice design, the following linear mixed model was used.

$$Y_{ijkl} = \mu + \text{Env}_i + \text{Rep}_j(\text{Env}_i) + \text{Block}_k(\text{Env}_i \text{ Rep}_j) + \text{Gen}_l + \text{Env}_i \times \text{Gen}_l + \varepsilon_{ijkl}$$

From the aforementioned equation, Y_{ijkl} represents the performance of the trait of interest, μ corresponds to the all-inclusive mean, Env_i represents the effect of the i th environment, $\text{Rep}_j(\text{Env}_i)$ represents the effect of the j th replication within the i th environment, $\text{Block}_k(\text{Env}_i \text{ Rep}_j)$ represents the effect of the k th incomplete block within the j th replication in the i th environment, Gen_l represents the effect of the l th genotype, $\text{Env}_i \times \text{Gen}_l$ represents the environment by genotype interaction, and ε_{ijkl} is the error variance. When calculating the BLUEs, genotypes and covariates were considered fixed effects of the model while other terms were included as random effects of the model. The covariate was considered as fixed effect of the model while all other terms were included in the random effects of the model to estimate the BLUPs. Heritability for the different traits was calculated as the ratio of the estimated genotypic variance to the estimated phenotypic variance ([Knapp et al., 1985](#)).

2.5 Genotyping and QTL mapping

The CML511×CML546 DH and parental lines were grown in a greenhouse. Maize leaf tissue samples were collected from young, healthy seedlings at the V3 stage (3–4 weeks old), stored at -80°C , and later freeze-dried for 72 h. High-quality genomic DNA was isolated from freeze-dried tissues using the standard CIMMYT laboratory protocol (CIMMYT, 2005). The DH lines were genotyped using DArTSeq in Canberra, Australia. Approximately 15,000 SNPs were used for further quality checks (Murithi et al., 2021). Trait Analysis by aSSociation Evolution and Linkage (TaSSEL) (Bradbury et al., 2007) was used to summarize genotype data by site, determine the allele frequencies, and implement quality screening. All SNP variants that were monomorphic between the two parents, had heterozygosity of >0.05 and a minor allele frequency of <0.05 , were filtered, and 1,264 high-quality SNPs were retained for QTL mapping.

Redundant markers were removed using the BIN tool in QTL IciMapping v.4.2 (Meng et al., 2015). In the parameter setting window for general information, eight functionalities were used to define the mapping population. In the indicator row, 1 was selected to denote QTL mapping in actual populations and 3 as the population type as this study used a DH population. Kosambi was set as the mapping function, marker position as the marker space type, 10 as the number of chromosomes, 230 as the size of the mapping population, and 6 as the number of traits.

The number of markers in each chromosome was specified in the chromosome information part. The scores for all the DArTseq markers were transformed into genotype codes following the scores of the parents (2 denoted the marker type of the first parent, 0 denoted the second parent, 1 for the F_1 marker type, and -1 for the missing markers) (Meng et al., 2015). The genetic linkage maps were constructed using the MAP functionality of QTL IciMapping v.4.2 (<http://www.isbreeding.net>). Three steps were followed in linkage map construction: grouping, ordering, and rippling. The logarithm of odds score was set at 3.0 for grouping. Ordering was performed using the ordering instruction with the nnTwo Opt algorithm. The sum of adjacent recombination frequencies (SARF) as the criterion and window size of 5 as the amplitude were used to ripple the marker sequence and to fine-tune the chromosome orders. All the outputting functionalities were checked, and the map was drawn using the MAP functionality (Meng et al., 2015).

In the phenotypic data, the BLUPs for the different traits were used as the input files for QTL identification across environments (Littell et al., 2007). The input file was loaded onto the project of IciMapping v.4.2 (Meng et al., 2015). In the parameter setting window ICIM-ADD, other parameters, such as step in scanning represented by cM and stepwise regression of phenotype on marker variables, were defined. An LOD threshold of 3.0 and 1000 permutations at $\alpha = 0.01$ were set to declare the significant QTL. The percentage of total phenotypic variance explained by an individual QTL was determined using stepwise regression. To ascertain the actual locations of the QTL for all the traits on the chromosomes, the physical position of the identified QTL was assigned based on the known physical position of the linked markers and also available at the maize genetics and genomics database (http://www.maizegdb.org/data_center/map), as described

by Berger et al. (2014). The individual QTLs were assigned names according to the QTL, trait name, chromosome, and marker position, as described in the work of Kibe et al. (2020a).

2.6 Genotyping and association mapping

The DNA of all 239 inbred lines of the association panel was extracted from seedlings at the 3–4 leaf stage and genotyped using the genotype-by-sequencing (GBS) platform at the Institute for Genomic Diversity, Cornell University, Ithaca, United States, using high-density markers, as per the procedure described by Elshire et al. (2011). SNP calling and imputation were conducted at Cornell University. For SNP calling, raw data in a FASTQ file together with the barcode information and Tags On Physical Map (TOPM) data, which had SNP position information, were used. We used TOPM data from AllZeaGBSv2.7 downloaded from Panzea (<https://www.panzea.org/>), which contained information for 955,690 SNPs mapped with B73 AGPv2 coordinates. SNP calling was then performed using the TASSEL-GBS pipeline (Glaubitz et al., 2014; Wang et al., 2020). Using TASSEL ver5.2 (Bradbury et al., 2007), SNPs with a heterozygosity of $>5\%$, MAF of >0.05 , and minimum count of 90% were excluded by filtering from raw GBS datasets, and 230,743 high-quality SNPs were retained for further analysis in the association panel. To explore the population structure and the ultimate number of subpopulations, principal component analysis (PCA) as described by Price et al. (2006) was conducted in TASSEL using SNPs across all panels. The first three principal components were instrumental to visualize the existing population stratification within the association panel, and this was clearly displayed in a 3D plot. The PCA plots of the association panel were computed using 230,743 SNPs; we then plotted the variance (y-axis) against the principal components (x-axis) to estimate the number of clusters within the population (Sanchez et al., 2018). The data point at which the increase in the number of principal components did not result in an increase in variance (leveling off) indicated the number of subgroups within the panel. To estimate the amount of genetic relatedness among individuals, a kinship matrix was explored.

GWAS was performed with different models to compare and choose the appropriate model with relatively less false positives. To detect marker–trait associations, GWAS was performed using the following models: (1) mixed linear models (MLMs; PCA + K + G) that incorporated PCA, kinship (K), and genotypic data as covariates; (2) the general linear model (GLM; PCA + G) which incorporated the genotype data (G) and the PCA (Q) that both acted as fixed effects to correct for the population structure; and (3) Fixed and random model Circulating Probability Unification (FarmCPU), in which the kinship (random) and the three-component analysis (fixed) were identified as covariates (Lipka et al., 2012). Single-locus GWAS models such as the GLM are characterized by high false positive rates, as a complementary model, and the MLM utilizes the Bonferroni correction to overcome the challenge of false positive rates and identify the loci of interest (Khan et al., 2021). The software TASSEL (Bradbury et al., 2007) was instrumental to run the GLM + PCA and MLM. The $-\log_{10} p$ values for all the analyzed SNPs for both GLS-DS and NCLB-DS were used to construct the Manhattan plots. Q–Q plots were plotted from the estimated $-\log_{10} (p)$ from the

association panel for GLS-DS and NCLB_DS traits. Analysis of the association panel was conducted in TASSEL based on 230,743 filtered SNPs. The R package 'FarmCPU' with the Genome Association and Prediction Integrated Tool (GAPIT) was used for GWAS analysis (Tang et al., 2016). The false discovery rate (FDR) was calculated for significant associations using the Benjamini and Hochberg (1995) correction method, with 8×10^{-5} as the threshold. To summarize GWAS results per chromosome, Manhattan scatter plots were generated by plotting genomic positions of the SNPs against their negative log base 10 of the p -values obtained from the GWAS model, with the F-test for the null hypothesis on the y-axis.

SNPs detected in the association panel were examined as polymorphisms in linkage disequilibrium with putative candidate genes from the B73 reference gene set (https://phytozome-next.jgi.doe.gov/jbrowse/index.html?data;Zea_mays_Zm-B73-REFERENCE-NAM-5.0.55) (Goodstein et al., 2012). Putative candidate genes were selected by delving into the information from Gene Ontology, Kyoto Encyclopedia of Genes and Genomes (KEGG), and protein families (Pfam) (Ashburner et al., 2000; Kanehisa and Goto, 2000; Bateman et al., 2004).

Genomic prediction was carried out with ridge regression BLUP (Zhao et al., 2012) within a biparental DH population for GLS, NCLB, and agronomic traits at five-fold cross-validation. The BLUEs across environments were used for the analysis. The same set of 1,264 high-quality uniformly distributed SNPs with no missing values and $MAF > 0.05$ were used. For the maize association panel, quality screening criteria of SNPs with $MAF > 0.10$ and no missing values were applied, and finally, 8,365 SNPs from the 230,743 SNPs were retained for the analyses. The prediction was 'within population', where training and validation sets were derived from within the biparental population. For each trait, 100 iterations were carried out for the sampling of the training and validation sets. The prediction accuracy was calculated as the correlation between the observed phenotypes and genomic estimated breeding values (GEBVs) divided by the square root of heritability (Dekkers, 2007).

3 Results

3.1 Phenotypic data

As expected for western Kenya, there was high natural disease pressure for both NCLB and GLS for all field trials of the biparental CML511×CML546 DH population (Figure 1A), as well as the association panel (Figure 1B). DS scores for both NCLB and GLS were highest at the final disease rating time point in all field trials, which corresponded in most cases to the last disease rating time point (Supplementary Table S1). A significant difference in resistance to NCLB was reported for the two parents CML511 and CML546 (p -value = 0.011831, $\alpha < 0.05$). Our data show that CML511 is moderately susceptible and CML546 is more resistant to NCLB. On the other hand, the two parents differed slightly but not significantly in resistance to GLS (p -value = 0.200588, $\alpha < 0.05$). For GLS, CML511 had an average score of 2.47 at the final rating, while CML546 had an average score of 4.0 at the final average DS score (Supplementary Figure S1). These

evaluations show that CML511 is resistant and CML546 is moderately susceptible to GLS. A large portion of the biparental population was extensively blighted by NCLB. Transgressive segregation was observed in the population for GLS, NCLB, and AD, as some of the genotypes were more resistant or susceptible compared to the parental lines in the biparental population (Figure 1).

The frequency distribution of GLS DS scores was fairly skewed toward resistance in the biparental population (Figure 1A). The frequency distribution of the DS scores for NCLB in the biparental population followed an approximately normal distribution, as shown in Figure 1A. The wide segregation of DS and AUDPC scores for NCLB provided more evidence for the quantitative control of resistance. The DH population also exhibited continuous distribution for the days to anthesis, days to silking, plant height, and ear height (Figure 1A). The GLS and NCLB DS scores for the maize association panel ranged from 1.5 to 6 (Figure 1B), which were similar to the biparental DH population scores, although the association panel trended toward higher GLS and lower NCLB scores. On the other hand, the use of the nine-point rating scale revealed extensive phenotypic variation in resistance to GLS and NCLB across the association panel, with the panel harboring more resistant lines (Figure 1B). The association panel was also characterized by shorter days to anthesis compared to the biparental population (Figure 1B).

3.2 Correlations between environments and variables

In the DH population, the correlation between environments for GLS DS was positive and highly significant ($p < 0.001$) (Supplementary Table S2). A moderately high correlation was observed between environments for NCLB DS scores (Supplementary Table S2). The correlation across environments for AD and SD was also highly significant at $p < 0.001$ (Supplementary Table S2).

The analyses of variance revealed significant genotypic and genotype × environment interaction variances for GLS, NCLB DS, and AUDPC values as well as other agronomic traits (Table 1). Heritability estimates on an entry mean basis were 0.81, 0.81, 0.79, and 0.80 (Table 1) for GLS DS, AUDPC for GLS, NCLB DS, and AUDPC for NCLB, respectively in the DH population. However, the heritability estimates for DS in the association mapping panel were lower (0.35 for GLS and 0.64 for NCLB).

Interestingly, the correlation analyses in the DH population showed low positive significant correlation between GLS (and AUDPC_GLS) and NCLB (Figure 2), indicating that there are different genomic loci that explain the variance for each disease. The agronomic traits for reproductive traits, namely, flowering time (AD) and days to silking (SD), were significantly negatively correlated with DS and AUDPC for both GLS and NCLB diseases (Figure 3). This indicated that maize lines with earlier maturity had higher DS. As expected, ear height (EH) was highly correlated with plant height (PH). GLS DS and AUDPC values were weakly correlated with PH and EH, whereas NCLB DS and AUDPC values were positively and significantly correlated with PH and EH

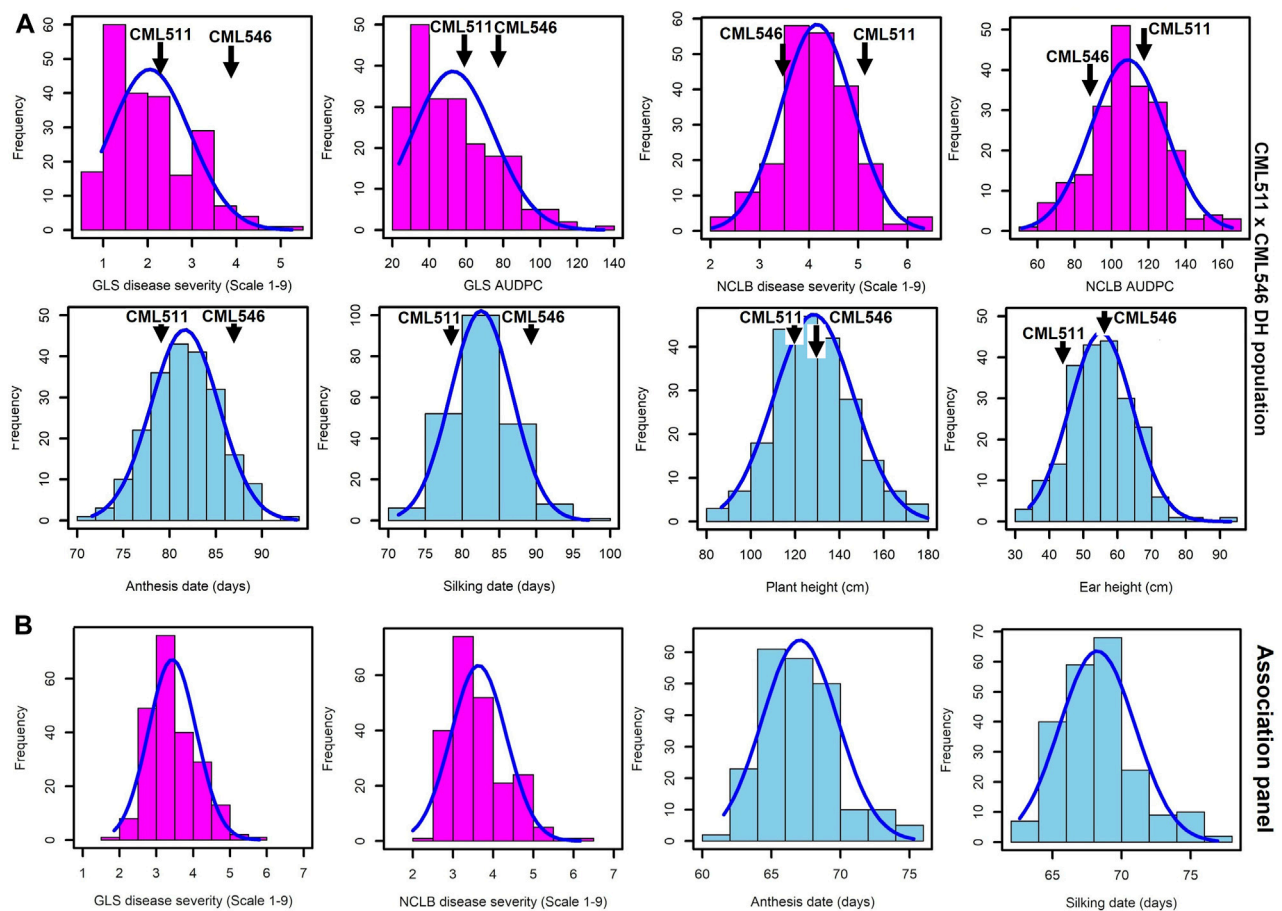


FIGURE 1

Frequency distributions for GLS, NCLB disease, and other agronomic traits, namely, anthesis date, silking date, plant height, and ear height, evaluated across the three locations in western Kenya. (A). Biparental CML511x CML546 DH population of 230 lines. (B). Association panel of 239 sub-tropical and tropical maize lines across the three locations. DS scores were for the last rating time point.

TABLE 1 Estimates of means, components of genotypic (σ^2_G), genotype \times environment interaction ($\sigma^2_{G \times E}$), error variances (σ^2_e), and heritability (h^2) for the biparental CML511x CML546 DH population and the association panel evaluated across three environments each for GLS, NCLB, and other agronomic traits.

	AD	SD	PH	EH	GLS	GLS	NCLB	NCLB	GLS ^a	NCLB ^a
					DS	AUDPC	DS	AUDPC		
Mean	81.74	82.69	127.89	55.02	2.36	52.27	4.58	108.46	3.11	3.62
σ^2_G	11.07**	14.37**	282.87**	67.79**	0.94**	382.44**	0.42**	311.71**	0.03*	0.07**
$\sigma^2_{G \times E}$	2.37**	3.63**	16.66**	5.94**	0.49**	213.07**	0.19*	156.61**	0.09**	0.04**
σ^2_e	11.59	12.6	219.57	101.49	0.52	175.63	0.45	216.91	0.16	0.17
h^2	0.84	0.85	0.90	0.84	0.81	0.81	0.79	0.80	0.35	0.64
LSD _{5%}	2.69	3.04	11.21	7.07	0.84	17.31	0.6	15.95	0.33	0.52
CV	4.17	4.29	11.59	18.31	30.56	25.36	14.73	13.58	34.09	18.06

AD, days to anthesis; SD, days to silking; PH, plant height; EH, ear height; DS, disease severity on a scale of 1–9; GLS, gray leaf spot; AUDPC, area under the disease progress curve; NCLB; northern corn leaf blight; CV, coefficient of variation; LSD, least significant difference; h^2 , broad-sense heritability; *, **significant at $p=0.05$ and 0.01 levels, respectively.

^aDisease severity scores of the maize association panel.

TABLE 2 QTL detected by integrated composite interval mapping analysis for resistance to GLS and NCLB in the DH population evaluated in multiple locations.

Trait name	QTL name ^a	Chr	Position (cM)	Left marker ^b	Right marker ^b	LOD	PVE (%)	TPVE (%)	Add	Fav parent
GLS DS	<i>qGLS1_54</i>	1	163	S1_283894617	S1_53456776	8.95	5.33	64.19	0.22	CML546
	<i>qGLS1_186</i>	1	372	S1_190286762	S1_185978658	21.86	15.17		-0.28	CML511
	<i>qGLS1_185</i>	1	383	S1_185978658	S1_143231392	11.34	9.01		0.21	CML546
	<i>qGLS2_30</i>	2	208	S2_30710232	S2_32668550	3.79	2.24		0.11	CML546
	<i>qGLS3_151</i>	3	92	S3_157562360	S3_150546157	5.31	3.37		0.13	CML546
	<i>qGLS5_07</i>	5	189	S5_7548544	S5_1579511	3.62	2.25		-0.11	CML511
	<i>qGLS5_16</i>	5	284	S5_15869219	S5_23093956	5.76	3.56		-0.14	CML511
	<i>qGLS7_158</i>	7	105	S7_158889984	S7_158892468	4.49	2.58		-0.11	CML511
	<i>qGLS9_129</i>	9	154	S9_135788881	S9_129671108	4.27	3.53		-0.13	CML511
	<i>qGLS10_50</i>	10	217	S10_43765534	S10_54916081	3.19	6.21		0.22	CML546
NCLB DS	<i>qNCLB1_230</i>	1	70	S1_229375633	S1_232878545	6.20	4.27	64.94	0.12	CML546
	<i>qNCLB1_303</i>	1	429	S1_303106691	S1_304299395	3.13	2.08		0.08	CML546
	<i>qNCLB2_220</i>	2	164	S2_223388206	S2_32056786	5.40	3.82		0.16	CML546
	<i>qNCLB3_02</i>	3	163	S3_2734515	S3_1173815	4.95	3.57		0.11	CML546
	<i>qNCLB3_50</i>	3	185	S3_65853211	S3_12761976	9.74	8.76		0.18	CML546
	<i>qNCLB4_200</i>	4	272	S4_200040593	S4_201402668	3.50	4.58		-0.12	CML511
	<i>qNCLB5_83</i>	5	80	S5_82971208	S5_160085856	3.60	4.56		0.13	CML546
	<i>qNCLB5_195</i>	5	113	S5_194106967	S5_198705622	3.60	2.76		0.10	CML546
	<i>qNCLB6_137</i>	6	204	S6_136207036	S6_137005821	3.21	2.17		-0.09	CML511
	<i>qNCLB6_153</i>	6	335	S6_151834390	S6_153165363	3.95	3.58		-0.20	CML511
	<i>qNCLB7_120</i>	7	96	S7_121214712	S7_47406965	7.51	5.58		-0.14	CML511
	<i>qNCLB7_174</i>	7	257	S7_170932028	S7_174093748	3.43	2.44		-0.10	CML511
	<i>qNCLB8_171</i>	8	276	S8_170418369	S8_171776990	7.85	5.77		0.14	CML546

GLS DS, gray leaf spot disease severity; NCLB DS; northern corn leaf blight disease severity; LOD, logarithm of odds; add, additive effect; PVE, phenotypic variance explained; fav parent, parental genotype from where a favorable allele is contributing.

^aQTL, name composed by the trait code followed by the chromosome number in which the QTL was mapped and a physical location of the QTL. QTL names are italicized.

^bThe exact physical position of the marker can be inferred from the marker's name, for example, S1_82702920: chromosome 1; 82,702,920 bp (Ref Gen_v3 of B73).

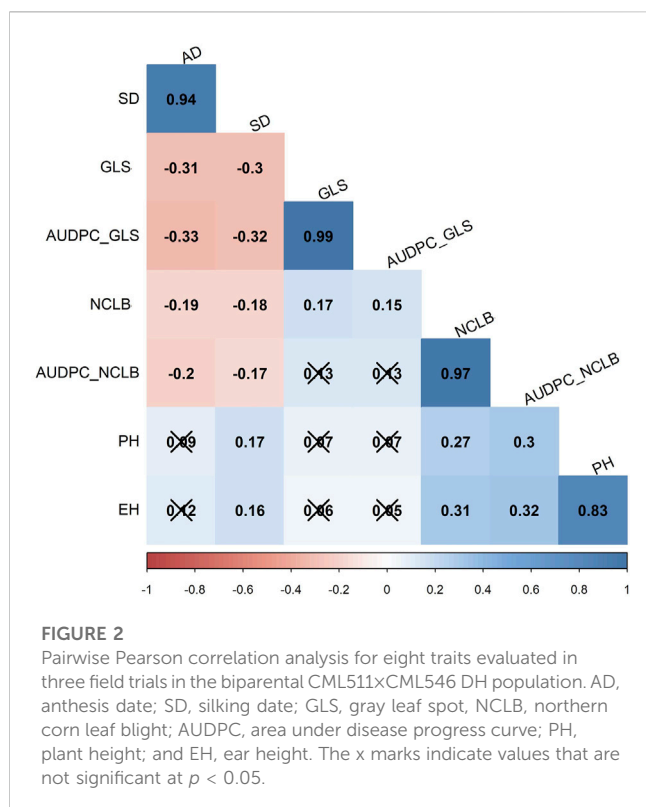
(Figure 3). There were weak positive and significant correlations between SD and PH/EH (Figure 3).

3.3 Construction of the genetic linkage map and QTL analyses

The linkage map for the CML511×CML546 DH population was constructed with a total of 1,264 SNP markers. The genetic linkage map spanned a total map length of 3,344.9 cM with 2.65 cM as the average distance between two adjacent markers. The linkage map, as shown in Supplementary Figure S2, covered most of the maize genome.

Several QTLs associated with resistance to GLS and NCLB with small additive effects were detected through inclusive composite interval mapping. QTL analyses revealed 10 QTLs distributed on chromosomes 1, 2, 3, 5, 7, 9, and 10 for GLS DS which individually explained 2.6%–15.2% of phenotypic variance and together

explained 64.19% of total phenotypic variance (Table 2). All 10 QTLs detected for GLS DS were also consistently detected for GLS AUDPC values (Supplementary Table S3). For NCLB DS, 13 QTLs distributed in all chromosomes except for chromosomes 9 and 10, individually explained 2.2%–8.7% which together contributed 64.94% of total phenotypic variation. AUDPC values for NCLB revealed nine QTLs together explained 45% of total phenotypic variance (Supplementary Table S3). Three QTLs on chromosomes 3, 6, and 7 were consistent across NCLB DS and AUDPC values. Among the agronomic traits, for AD, nine QTLs detected together explained 63% of the total phenotypic variance, and one QTL on chromosome 8 (*qAD8_137*) was a major effect QTL which explained 15.84% of phenotypic variance ((Supplementary Table S4). For SD, six QTLs were detected which together explained 60% of the total phenotypic variance. There were three QTLs on chromosomes 1, 4, and 8 that were consistent for both AD and SD. For PH, four QTLs were identified which together explained 52% of total phenotypic variance. One major effect QTL (*qPH8_129*)



detected on chromosome 8 explained 22% of phenotypic variance. For EH, six minor effect QTLs and one major effect (*qEH8_128*) QTL were detected which together explained 56% of total phenotypic variance. QTL mapping predominantly revealed that the additive gene effect defined the gene action for resistance to GLS and NCLB.

3.4 GWAS analysis

After a quality check of GBS SNP markers, 230,743 SNPs were retained for the final association analyses (Supplementary Figure S3). The kinship matrix for these 239 lines was projected in the form of a heatmap which shows the magnitude of the relationship between the individuals (Supplementary Figure S4). This clearly showed that there is no strong population structure in the association panel used here. PCAs revealed five subpopulations within the panel (Supplementary Figure S5).

Association analyses for GLS DS and NCLB DS data were performed with GLM, MLM, and FarmCPU models (Figure 3). For both GLS and NCLB DS traits, for the GLM model, the observed p -values showed higher deviation from the expected p -values which may cause high false positives. On the other hand, for the MLM model, though the observed p -values were closer to the expected p -values, overfitting of the model is observed. For the FarmCPU model, the observed p -values were close to the expected p values and were more effective in controlling the false positives (Figure 3). The FarmCPU model is known to use both fixed and random effects models iteratively which helps in avoiding overfitting of the model by stepwise regression (Liu et al., 2016). Therefore, in this study, we used the FarmCPU model in the association mapping.

Association analyses revealed 11 and 18 SNPs significantly associated with GLS DS and NCLB DS, respectively (Figure 3; Table 3). For GLS-DS, the identified SNPs were distributed across all chromosomes except for chromosomes 3, 6, and 7 (Figure 3). The Manhattan plot revealed that for GLS DS, the highest peak was reported on chromosomes 4 and 8 (Figure 3A), while for NCLB DS, the highest peak was reported on chromosome 8 (Figure 3C). Furthermore, we determined whether any of these significant GLS or NCLB disease-associated SNPs identified in the association analysis is co-located with the QTL for GLS or NCLB in the biparental DH population. Two SNPs on chromosome 1 (*S1_192041854* and *S1_253381765*) were co-located with *qGLS1_54* detected for both GLS DS and AUDPC values in the DH population (Table 2; Table 3). Another SNP on chromosome 9 (*S9_130213878*) was found to be collocated within the *qGLS9_129* and *qG_AUDPC9_129* QTL region (Table 2; Table 3). Among the 16 SNPs identified for NCLB, marker *S5_83980678* is found to be within the region of NCLB DS QTL *qNCLB5_83* (Table 2; Table 3).

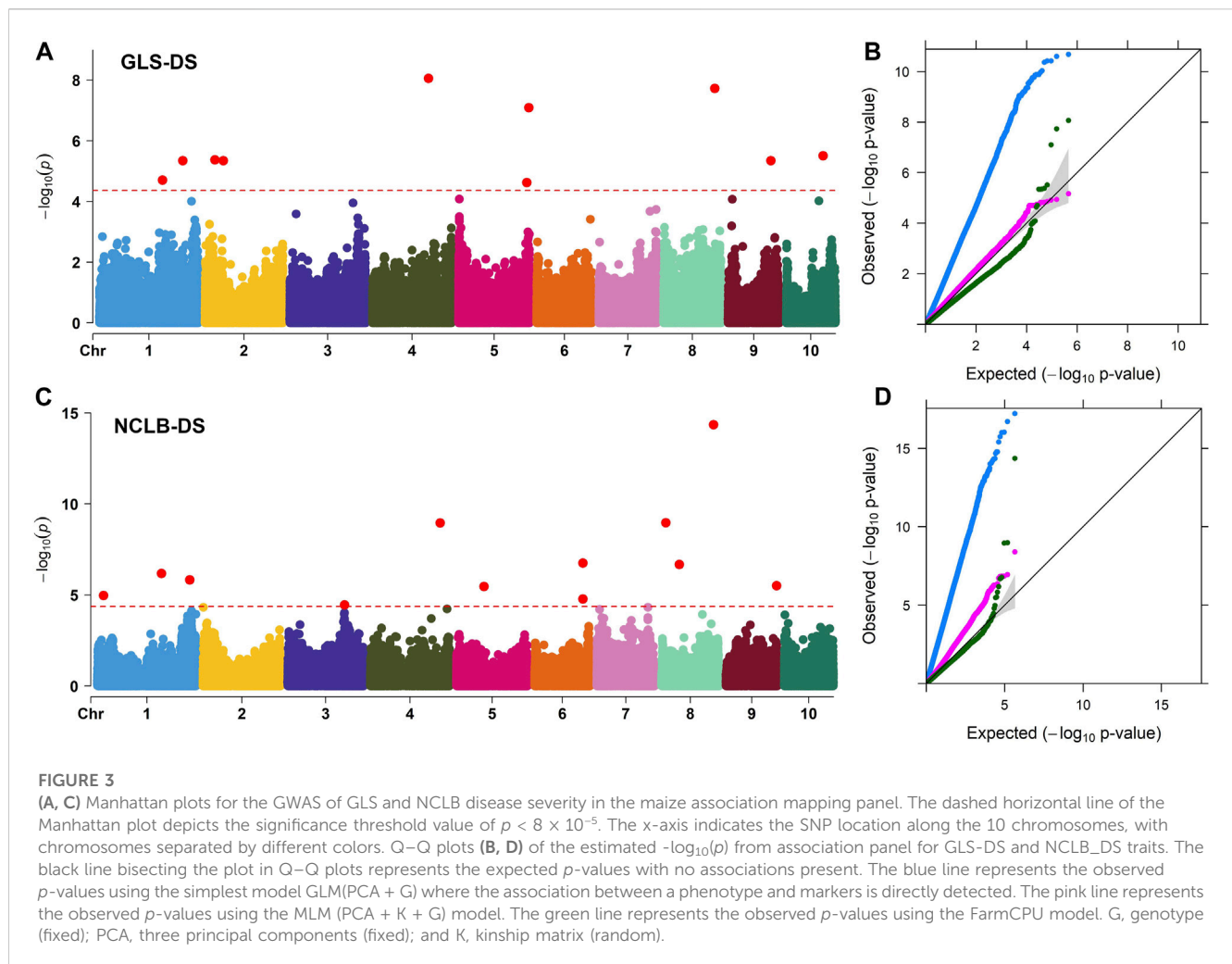
To elucidate the molecular and physiological mechanisms controlling GLS and NCLB DS, candidate genes were identified. On all chromosomes, a total of 24 candidate genes were discovered (Table 3). Four candidate genes closely associated with the SNPs for GLS resistance were identified, namely, *Zm00001eb077270*, *Zm00001eb034870*, *Zm00001d017831*, and *Zm00001eb211960* (Table 3). There were six candidate genes with defense response annotations that were associated with SNPs for NCLB resistance (*Zm00001eb201110*, *Zm00001eb035640*, *Zm00001eb232660*, *Zm00001eb285080*, *Zm00001eb144960*, and *Zm00001d001787*) (Table 3).

Genomic prediction captures all variations from small to large effects, which helps in improving complex traits. Prediction correlations obtained from cross-validations are commonly used to know the effectiveness of genomic predictions for different traits. In this study, we applied genomic predictions within the DH population and association panel for disease traits and also for agronomic traits (Figure 4).

As expected, the mean prediction correlations for the DS traits were higher in the DH population (GLS DS = 0.75, NCLB DS = 0.68) than those in the association panel (GLS DS = 0.63, NCLB = 0.49) (Figure 4). This was because of the highly relatedness among the DH lines compared to the lines in the association panel. In addition, relatively high-average prediction correlations were obtained for the other traits when validated in the DH population, namely, 0.73, 0.66, 0.71, 0.72, 0.75, and 0.71 for GLS-AUDPC, NCLB-AUDPC, AD, SD, PH, and EH, respectively.

4 Discussion

GLS and NCLB are economically important foliar diseases of maize. Understanding their genetic basis of resistance is valuable to designing an effective breeding strategy (Beyene and Prasanna, 2020). The DH population used in this study was artificially inoculated with GLS; however, the locations evaluated were also a hotspot for NCLB, so we observed both GLS and NCLB disease symptoms in the same population. For early scoring, symptoms for both diseases were clearly distinguishable, which helped to score the data with more accuracy. Scoring at a later stage of disease



development was tricky due to bigger blight merging with leaf spots, so for the analyses, we used the third DS score for both GLS and NCLB. The association mapping panel was also evaluated in natural hot spots for GLS and NCLB, and scoring was performed only once at a grain-filling stage when clearly distinguished GLS and NCLB symptoms were observed. Therefore, the collected DS data represent the real response of these lines to the respective diseases. Nevertheless, both diseases appearing at the same growth stage and in the same experiment can lead to some confounding effects. Most of the lines in both the DH population and the association panel fall into the categories of resistant and moderately resistant, with a few in moderately susceptible but none in the completely susceptible category (Figure 1). Overall, the phenotypic data in this study showed a normal distribution for NCLB and GLS DS scores in both the biparental CML511×CML546 DH population and the association panel which supports the quantitative nature of resistance in these diseases (Nyanapah et al., 2020). The parental line CML511 exhibited a moderate level of resistance to GLS congruent with the observations of Kibe et al. (2020a). We also observed significant genotypic and genotype × environment interaction variance and moderate-to-high heritabilities in both the DH population and the association panel, indicating good prospects for introgressing GLS and NCLB resistance in breeding

programs. Observed heritability estimates for GLS, NCLB, and other agronomic traits in the DH population are consistent with earlier studies (Wisser et al., 2011; Benson et al., 2015).

4.1 Trait correlations

There were moderate correlations in GLS and NCLB DS scores for the biparental DH population between the field trials, indicating that trait expression was relatively consistent between the evaluated locations (Supplementary Table S2). On the other hand, a significantly negative correlation was observed between DS data of GLS and NCLB, with the flowering time traits AD and SD in this study (Figure 2). This implies that lower values for AUDPC (implying higher levels of disease resistance) corresponded with longer AD or SD. Such negative correlations have also been reported in other studies (Asea et al., 2009; Wisser et al., 2011; Kolkman et al., 2020). On the contrary, some studies reported a positive correlation between GLS resistance and flowering time (Balint-Kurti et al., 2008; Zwonitzer et al., 2010; Benson et al., 2015; Mammadov et al., 2015; Liu et al., 2016). This suggests the cautious use of flowering time in the selection of lines for resistance to GLS and NCLB.

TABLE 3 Chromosomal position and SNPs significantly associated with GLS disease severity (GLS_DS) and northern corn leaf blight disease severity (NCLB-DS) detected by SNP-based GWAS in the association mapping panel.

SNP ^a	Chr	MLM <i>p</i> -value	MAF ^b	H&B ^c <i>p</i> -value	Effect	Candidate gene	Gene annotation
GLS-DS							
S4_170027069	4	8.68E-09	0.35	0.00	0.16	<i>Zm00001eb189650</i>	K13120—protein FAM32A (FAM32A)
S8_155438805	8	1.86E-08	0.47	0.00	−0.20	<i>Zm00001eb360540</i>	Cation efflux protein
S5_214099678	5	7.99E-08	0.29	0.01	0.18	<i>Zm00001eb254100</i>	Zinc finger FYVE domain-containing protein
S10_112359288	10	3.09E-06	0.13	0.13	−0.20	<i>Zm00001eb421180</i>	Copper transport protein atox1-related (abiotic stress)
S2_29666484	2	4.25E-06	0.38	0.13	−0.13	<i>Zm00001eb077270</i>	Wall-associated receptor kinase galacturonan-binding domain (defense)
S1_253381765	1	4.51E-06	0.40	0.13	−0.13	<i>Zm00001eb049550</i>	Sel-1-like protein
S9_130213878	9	4.53E-06	0.10	0.13	−0.24	<i>Zm00001eb393380</i>	No associated annotations
S2_55483916	2	4.53E-06	0.26	0.13	0.13	<i>ZM00001EB083120</i>	No associated annotations
S1_192041854	1	1.98E-05	0.26	0.51	0.15	<i>Zm00001eb034870</i>	DNA damage–repair/tolerance protein (plant defense)
S5_208091867	5	2.37E-05	0.21	0.55	0.12	<i>Zm00001eb251710</i>	Brevis radix domain/regulator of chromosome condensation (plant defense)
S5_2923669	5	8.36E-05	0.24	1.00	0.13	<i>Zm00001eb211960</i>	NAD-dependent epimerase/dehydratase//cinnamoyl-CoA reductase-like
NCLB-DS							
S8_157881780	8	4.41E-15	0.15	0.00	−0.57	<i>GRMZM2G059590</i>	Uncharacterized protein LOC103636219
S8_12990030	8	1.07E-09	0.16	0.00	−0.27	<i>Zm00001d008560</i>	No associated annotations
S4_212315234	4	1.11E-09	0.16	0.00	0.25	<i>Zm00001eb201110</i>	ATP-binding cassette transporter//ABC transporter (plant defense)
S6_147054037	6	1.75E-07	0.08	0.01	0.25	<i>Zm00001eb285130</i>	TPR repeat-containing THIOREDOXIN TDX
S8_54094984	8	2.11E-07	0.49	0.01	−0.17	<i>Zm00001eb341620</i>	Thioesterase superfamily member-related
S1_194762510	1	6.65E-07	0.22	0.03	−0.26	<i>Zm00001eb035640</i>	AUX/IAA protein//B3 DNA-binding domain//auxin response factor//DNA-binding pseudobarrel domain (plant defense)
S1_280826386	1	1.48E-06	0.28	0.05	0.17	<i>Zm00001d034003</i>	Seed maturation family protein
S9_153843575	9	3.08E-06	0.45	0.09	0.16	<i>Zm00001eb400490</i>	Pre-mRNA-processing factor
S5_83980678	5	3.39E-06	0.14	0.09	0.22	<i>Zm00001eb232660</i>	Helicase superfamily/ATP-binding domain (plant defense)
S1_18630129	1	1.05E-05	0.45	0.24	−0.13	<i>Zm00001eb006570</i>	WD and tetratricopeptide repeats protein 1 (WDTIC1)
S6_146813774	6	1.64E-05	0.06	0.34	−0.21	<i>Zm00001eb285080</i>	Protein kinase domain (plant defense)
S3_173387708	3	3.49E-05	0.09	0.67	−0.21	<i>Zm00001eb144960</i>	Lipoxygenase (plant defense)
S2_923555	2	4.77E-05	0.32	0.79	−0.16	<i>Zm00001d001787</i>	Cleavage and polyadenylation specificity factor subunit 5 (plant defense)
S7_155701108	7	4.77E-05	0.14	0.79	0.21	<i>Zm00001d021552</i>	Protein of unknown function
S4_233626821	4	5.87E-05	0.47	0.88	−0.15	<i>Zm00001eb204230</i>	Voltage- and ligand-gated potassium channel
S7_8047716	7	6.13E-05	0.24	0.88	−0.12	<i>Zm00001d018877</i>	Plastocyanin-like domain (Cu_bind_like)

^aThe exact physical position of the SNP can be inferred from the marker's name, for example, S5_51353429: chromosome 5; 51353429 bp (Ref Gen_v2 of B73).

^bMinor allele frequency. Candidate gene names are italicized.

^cFalse discovery rate calculated by using the Benjamini and Hochberg correction method.

In Kenya and Uganda, the main maize-growing area is frequently affected by GLS and NCLB pathogens (Borchardt et al., 1998; Nsibo et al., 2021). When both pathogens affect the maize at the same time, more pronounced necrotic symptoms are

the major concern which are probably due to the synergistic interactions of both pathogens. However, a weak correlation was observed between GLS DS and NCLB DS. One of the QTLs identified for GLS (*qGLS2_30*) was in proximity with the QTL

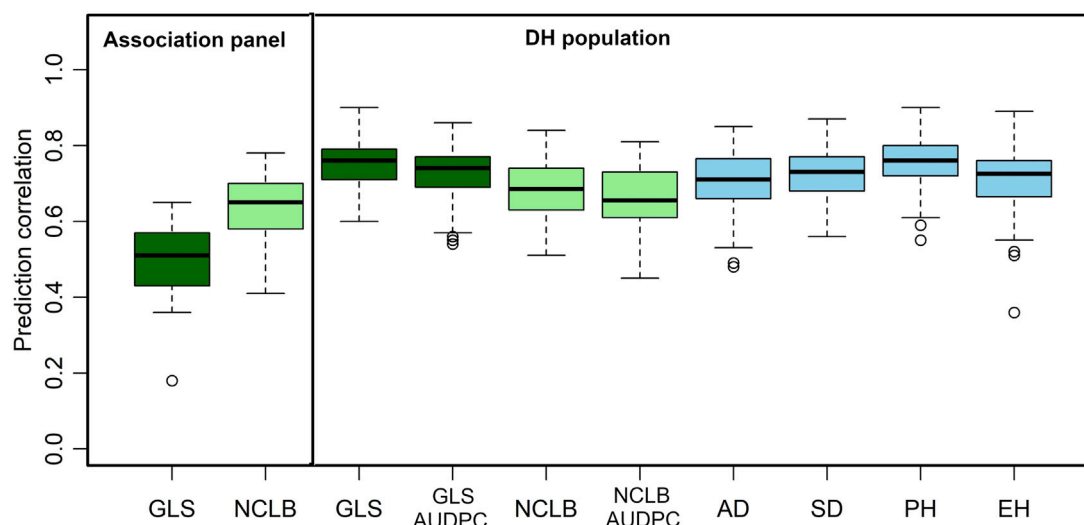


FIGURE 4

Box-whisker plots for the accuracy of genomic predictions assessed by five-fold cross-validation within association and DH population. AD, days to anthesis; SD, days to silking; PH, plant height; EH, ear height; GLS, gray leaf spot; AUDPC, area under the disease progress curve; NCLB, northern corn leaf blight.

identified for NCLB (*qNCLB2_220*). In the comparison of SNPs associated with GLS and NCLB DS in association mapping, it was observed that two SNPs (*S1_194762510* and *S1280826386*) associated with NCLB DS were collocated within GLS QTL (*qGLS1_54*) (Tables 2, 3). This suggests that there are some common regions contributing to resistance for both diseases. On the other hand, the observed weak correlation between the DS of the two diseases could be attributed to the different infection strategies of the associated pathogens. *Cercospora zeina* is an apoplastic necrotroph and a hemibiotroph, while *E. turcicum* is apoplastic but then enters the vascular system of the leaf (Kotze et al., 2019). They also exploit different pathogenicity factors in causing disease symptoms (Swart et al., 2017; Human et al., 2020).

4.2 QTLs associated with GLS resistance

Most of the QTLs detected for GLS DS were also detected for GLS AUDPC (Table 2). This was well supported by the observed strong correlation ($r = 0.99$) between GLS DS and GLS AUDPC (Figure 2). A major QTL for GLS resistance (DS and AUDPC), *qGLS1_186*, which explained 15.16% of the phenotypic variance, overlapped with *qGLS1_185* which also explained 9.01% of the phenotypic variance. Intriguingly, this major QTL has favorable alleles from the donor parent CML511. Kibe et al. (2020a) using the CML550×CML511 DH population also detected major effect QTL *qGLS1-155* which was located within the physical position of 154–157 Mbp which overlapped with the QTL detected in the present study (*qGLS1_185*) spanning between 143 and 185 Mbp. Sun et al. (2021) also fine-mapped a major effect QTL at the 187–189 Mb region, and the reported flanking markers would be useful to validate in tropical germplasm.

The chromosome bin 1.06 was described as a QTL hotspot for GLS resistance as many studies reported earlier (Lehmensiek et al.,

2001; Shi et al., 2007; Balint-Kurti et al., 2008; Berger et al., 2014; He et al., 2018; Lopez-Zuniga et al., 2019; Sun et al., 2021). The chromosome bin 1.06 also harbors resistance genes to several other diseases like common rust, southern leaf blight (SLB), ear rot, and NCLB (Freyemark et al., 1993; Wisser et al., 2006; Chung et al., 2010b; Zwonitzer et al., 2010; Poland et al., 2011; Jamann et al., 2015). Chung et al. (2010b) demonstrated that the NCLB resistance QTL at bin 1.06 was important to protect the host against fungal penetration of *E. turcicum* using an introgression line population. The chromosomal region has also been associated with effects on diverse traits such as grain yield and its components, anthesis silking interval, and root and shoot traits under both water stress and optimal water environments (Ribaut et al., 1996; Tuberosa et al., 2002; Landi et al., 2010). The concentrated mapping of QTL for several traits, including multiple disease resistance in this chromosomal region, provides breeders and geneticists an opportunity to dissect them further and find tightly linked flanking markers so that this region can be utilized to develop cultivars with multiple disease resistance.

The *qGLS2_30* QTL identified in this study in the chromosomal bin 2.04 overlapped with QTL reported in earlier studies using different mapping populations (Balint-Kurti et al., 2008; Lennon et al., 2017). The QTL *qGLS3_151* is placed between 150 and 157 Mbp in the chromosomal bin 3.05, which has previously been identified as conditioning resistance to SLB and GLS (Zwonitzer et al., 2010; Kump et al., 2011). The QTL *qGLS7_158* positioned between 157 and 159 Mbp in the chromosomal bin 7.04 was also previously reported by Berger et al. (2014) for GLS resistance. Another QTL *qGLS5_16* in the chromosomal bin 5.03 is also known to have several reported markers for GLS resistance in a number of association mapping studies (Bubeck et al., 1993; Clements et al., 2000; Lehmensiek et al., 2001; Shi et al., 2007; Zhang et al., 2012; Benson et al., 2015). Overall, many of the QTLs detected in the present study overlapped between the biparental

CML511 × CML546 DH population and the association panel as well as earlier studies (Supplementary Table S5). The major effect QTL for both GLS on chromosome 1 is of immediate interest to be used in resistance breeding.

Among the nine candidate genes identified for GLS resistance through association mapping, one on chromosome 2 (*Zm00001eb077270*) encodes a putative receptor-like protein kinase which are transmembrane signaling proteins that are able to sense changes in the extracellular environment such as pathogen invasion (Decreux et al., 2006; Qi et al., 2023). Another candidate gene on chromosome 1 (*Zm00001eb034870*) encodes DNA damage–repair/tolerance protein that harbors a leucine-rich repeat domain which serve as the first line of defense in response to pathogen-associated molecular patterns (Ng and Xavier, 2011). The candidate gene *Zm00001eb144960* on chromosome 3 encodes lipoxygenases which are known to be associated with pest resistance, response to wounding, and plant defense mechanisms where it was reported to be involved in the early response to pathogen attack (Peng et al., 1994). Overall, the detected candidate genes in the association study have annotations inferring direct or indirect involvement in plant defense.

4.3 QTL and SNPs associated with resistance to NCLB

This study identified 13 QTLs for NCLB DS and nine QTLs for NCLB AUDPC. Three of these QTLs were common for both the DS and AUDPC NCLB traits. The first example of this was QTLs *qNCLB3_50* and *qN_AUDPC3_50* that co-localized in the same position in bin 3.04. This significant QTL for NCLB resistance has favorable alleles from the parent CML546 (the more resistant parent). Previous research reports also identified bin 3.04 as a QTL hotspot conditioning resistance to multiple diseases including NCLB, SLB, and GLS (Lehmensiek et al., 2001; Wisser et al., 2006; Shi et al., 2007; Zwonitzer et al., 2010; Kump et al., 2011; Liu et al., 2016; Lennon et al., 2017; Martins et al., 2019).

The QTL *qNCLB5_83* was positioned in the chromosomal bin 5.04. According to Miedaner et al. (2020), up to eight QTLs have been localized in this bin showing the importance of this region for NCLB resistance. Interestingly, the SNP identified through association mapping (*S5_83980678*) is positioned within this QTL region (Table 2, 3). It is associated with a candidate gene (*Zm00001eb232660*) that encodes a DNA helicase/ATP-binding domain. This type of domain has a catalytic function in unwinding of the double-stranded DNA that is instrumental in the repair of damaged DNA and DNA replication (Koonin, 1993).

Similarly, *qNCLB6_153* for NCLB DS overlapped with *qN_AUDPC6_153* for the AUDPC on chromosome 6 (bin 6.05). Up to four QTLs were reported from different studies in the same region (Miedaner et al., 2020). Another pair of QTL, *qNCLB8_171* and *qN_AUDPC8_171*, corresponded with a previously reported NCLB resistance QTL by Galiano-Carneiro et al. (2021). Interestingly, none of the NCLB DS QTLs detected in this study were found in the same position with chromosomal

bins associated with the qualitative *Ht* genes (Galiano-Carneiro and Miedaner, 2017).

An association study on NCLB revealed a significant marker linked to a candidate gene *Zm00001eb201110* on chromosome 4 which encodes for an ATP-binding cassette (ABC) transporter. Plant proteins with this function are known to be associated with resistance to fungal and bacterial pathogens through the transmembrane transport of jasmonic acid or antimicrobial secondary metabolites (Zhang et al., 2020). Using GWAS, many studies showed the association of ABC transporter genes with NCLB resistance (Poland et al., 2011; Ding et al., 2015). Another candidate gene, *Zm00001eb285080*, on chromosome 6 encodes a protein kinase, a function known to be important in regulating the response of plants to pathogen attack (Lehti-Shiu and Shiu, 2012). There is strong evidence that protein kinases play a pivotal role in resistance to NCLB (Poland et al., 2011; Ding et al., 2015; Kolkman et al., 2020).

4.4 QTLs associated with agronomic data

The major QTL for flowering time *qAD8_137* was collocated with QTLs *qSD8_137* for SD, *qPH8_129* for PH, and *qEH8_128* for EH (Supplementary Table S4). These QTLs also explained the major effect of phenotypic variance of 15.8%, 21.4%, 22.39%, and 22.98% for AD, SD, PH, and EH, respectively. Several studies also recognized chromosomal bin 8.05 as a hotspot for flowering time QTL and genes (Balint-Kurti et al., 2008; Buckler et al., 2009; Van Inghelandt et al., 2012). Interestingly, two qualitative resistance genes, *Ht2* and *Htn1*, were also detected on the same chromosomal bin 8.05 (Galiano-Carneiro and Miedaner, 2017; Hurni et al., 2015). The genetic mechanisms underlying flowering time in this study were largely characterized by additive gene action. These results agree with the findings of Buckler et al. (2009) who reported that variations in days to flowering are due to the joint effect of many minor QTLs with additive effect.

Intriguingly, some of the QTLs associated with flowering time overlapped with the NCLB and GLS resistance QTL. For instance, *qAD1_60* for flowering time shared the same flanking markers as *qGLS1_54* (Table 2; Supplementary Table S4), and the two SNPs (*S1_192041854*, *S1_253381765*) for GLS DS and two SNPs (*S1_194762510*, *S1_2800826386*) for NCLB DS detected through association mapping are also positioned in this region. Another QTL *qGLS9_129* also had the same flanking markers as *qAD9_130*, and one SNP from association mapping (*S9_130213878*) was also detected in the same region for GLS DS. The NCLB QTL *qNCLB1_230* overlapped with *qAD1_227* on the maize chromosomal bin 1.07 (Table 2; Supplementary Table S4). The QTLs *qN_AUDPC2_188* and *qPH2_176* overlapped on chromosomal bin 2.06, sharing the flanking markers. This was further supported by the positive correlation between PH and NCLB AUDPC (Figure 2). On the other hand, Galiano-Carneiro et al. (2021) reported a negative correlation between PH and NCLB DS. There were no common QTL regions identified for PH and EH that spanned the same chromosomal regions as GLS DS and AUDPC. This is supported by the observed negative correlation between the two traits (Figure 3).

4.5 Genomic prediction of disease and agronomic traits

Compared to the association panel, the high-prediction correlations in the DH population for GLS and NCLB could be attributed to higher similarity or relatedness of individuals between the training set and the prediction set (Figure 4; Lorenz and Smith, 2015). GLS DS and GLS AUDPC exhibited slightly higher prediction accuracies compared to NCLB DS and other agronomic traits. Agronomic traits, such as AD, SD, PH, and EH, are characterized by more complex genetic networks, under the control of numerous QTLs, and affected by the influence of the environment (Wallace et al., 2014). This presents a challenge in improving them through phenotypic selection (Du et al., 2021). Kibe et al. (2020a) reported low-to-moderate prediction correlations within populations and high values when different related populations were combined and used in prediction. Similarly, Galiano-Carneiro et al. (2021) reported prediction accuracies of moderate levels (0.55) for prediction within families. Technow et al. (2013) reported prediction accuracies of 0.58 and 0.55 when using a small population size of 75 individuals. It has been reported that there is no difference among hybrids advanced through the genomic selection or phenotypic selection in their response to NCLB and GLS, with the genomic selection being relatively cheaper than the phenotypic selection (Beyene et al., 2019; 2021). Overall, the use of genomic selection has potential to improve the resistance to GLS and NCLB in breeding populations and could lead to the development of multiple disease-resistant lines and hybrids.

5 Conclusion

GLS and NCLB are the major biotic stresses that hinder maize production in high-yielding maize-growing areas in East Africa, such as western Kenya. The use of genomic tools can provide useful information to fast track the development of disease-resistant varieties. In this study, we aimed to identify and validate genomic regions associated with GLS and NCLB resistance in biparental and association mapping populations evaluated in multiple locations in western Kenya. We identified 10 and 11 QTLs for GLS resistance and 18 and 16 QTLs for NCLB resistance in the DH population and association mapping population, respectively. We detected a major QTL for GLS resistance, *qGLS1_186*, which explained 15.2% phenotypic variance and *qNCLB3_50* for NCLB resistance, explaining 8.8% of the phenotypic variance. Several common QTL regions between linkage mapping and association mapping and between NCLB and GLS AUDPC traits were detected. A negative correlation between flowering time and severity of the two diseases was reported. Several QTLs identified in the present study were also co-localized with the QTL previously mapped for GLS and NCLB resistance. Our study highlights that the combined use of linkage mapping and genomic selection is an effective strategy for the improvement of resistance. Genomic prediction sheds light on new ways to improve breeding for disease resistance with optimum allocation of resources and lays the foundation for a new era of resistance breeding.

Data availability statement

The datasets presented in this study can be found in online repositories: data.cimmyt.org/dataset.xhtml?persistentId=hdl:11529/10548956 and zenodo.org/records/10046213.

Author contributions

DO: conceptualization, data curation, formal analysis, validation, and writing—original draft. MD: conceptualization, funding acquisition, project administration, resources, supervision, and writing—review and editing. DB: conceptualization, methodology, resources, supervision, and writing—review and editing. YB: funding acquisition, investigation, project administration, supervision, and writing—review and editing. DN: conceptualization, formal analysis, methodology, and writing—review and editing. CJ: data curation, formal analysis, methodology, and writing—review and editing. SM: investigation and writing—review and editing. MG: data curation, formal analysis, methodology, software, visualization, writing—original draft, and writing—review and editing.

Funding

The authors declare financial support was received for the research, authorship, and/or publication of this article. This study was supported by the Kenya–South Africa joint science and technology research collaboration. The grants received from the National Research Fund, Kenya, and the National Research Foundation (NRF), South Africa (grant # 105806), toward this research are hereby acknowledged. This study was also supported by CIMMYT–Nairobi. CIMMYT received support from the United States Agency for International Development, Foundation for Food and Agriculture Research (FFAR), and the Bill and Melinda Gates Foundation (BMGF) under AG2MW (Accelerating Genetic Gains in Maize and Wheat for Improved Livelihoods, B&MGF Investment ID INV-003439) project, the CGIAR Research Program on MAIZE.

Acknowledgments

The authors thank the CIMMYT field and laboratory technicians for phenotypic evaluations and sample preparation for genotyping. They are also grateful to the Buckler Lab at Cornell University, United States, for genotyping the maize populations and providing the marker information and Diversity Arrays Technology (DArT), Canberra, Australia for genotyping DH population. Part of the data in this manuscript was from the thesis work conducted by DO.

Conflict of interest

Authors DO and CJ were employed by Crop Science Division Bayer East Africa Limited.

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.

The authors declare that they were editorial board members 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/fgene.2023.1282673/full#supplementary-material>

References

- Alvarado, G., López, M., Vargas, M., Pacheco, A., Rodríguez, F., Burgueño, J., et al. (2015). META-R (multi environment trial analysis with R for windows) version 5.0 - CIMMYT research software dataverse - CIMMYT dataverse network. Available at: <https://data.cimmyt.org/dataset.xhtml?persistentId=doi:10.11529/10201> (Accessed April 15, 2021).
- Asea, G., Vivek, B. S., Bigirwa, G., Lipps, P. E., and Pratt, R. C. (2009). Validation of consensus quantitative trait loci associated with resistance to multiple foliar pathogens of maize. *Phytopathology* 99 (5), 540–547. doi:10.1094/phyto-99-5-0540
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* 25 (1), 25–29. doi:10.1038/75556
- Balint-Kurti, P. J., Wissner, R. J., and Zwonitzer, J. C. (2008). Use of an advanced intercross line population for precise mapping of quantitative trait loci for gray leaf spot resistance in maize. *Crop Sci.* 48, 1696–1704. doi:10.2135/cropsci2007.12.0679
- Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S., et al. (2004). The Pfam protein families database. *Nucleic Acids Res.* 32, D138–D141. Database issue. doi:10.1093/nar/gkh121
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 57, 289–300. doi:10.1111/j.2517-6161.1995.tb02031.x
- 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 (3), 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
- Berger, D. K., Mokgobu, T., de Ridder, K., Christie, N., and Aveling Theresa, A. S. (2020). Benefits of maize resistance breeding and chemical control against northern leaf blight in smallholder farms in South Africa. *South Afr. J. Sci.* 116 (11–12), 113–119. doi:10.17159/sajs.2020/8286
- 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, 1502. 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, 685488. doi:10.3389/fpls.2021.685488
- Beyene, Y., and Prasanna, B. M. (2020). Accelerating genetic gains in maize and wheat: genetic gains in CIMMYT maize breeding program in Africa. Available at <https://excellenceinbreeding.org/sites/default/files/u1025/CIMMYT%20Maize%20Genetic%20Gains%20in%20Africa%20%28AGG-EiB%20Workshop%3B%20Nov%2018%202020%29.pdf> (Accessed August 20, 2021).
- Borchardt, D. S., Welz, H. G., and Geiger, H. H. (1998). Genetic structure of *Setosphaeria turcica* populations in tropical and temperate climates. *Phytopathology* 88 4, 322–329. doi:10.1094/phyto.1998.88.4.322
- Bradbury, J. P., 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 (19), 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 (4), 838–847. doi:10.2135/cropsci1993.0011183x00300040041x
- Buckler, E. S., Holland, J. B., Bradbury, P. J., Acharya, C. B., Brown, P. J., Browne, C., et al. (2009). The genetic architecture of maize flowering time. *Science* 325 (5941), 714–718. doi:10.1126/science.1174276
- Challa, S., and Neelapu, N. R. R. (2018). "Chapter 9 - genome-wide association studies (GWAS) for abiotic stress tolerance in plants," in *Biochemical, physiological and molecular avenues for combating abiotic stress tolerance in plants*. Editor S. H. Wani (Academic Press), 135–150. doi:10.1016/b978-0-12-813066-7.00009-7
- Chen, G., Wang, X., Long, S., Jaqueth, J., Li, B., Yan, J., et al. (2015). An update on inflammation in the acute phase of intracerebral hemorrhage. *Mol. Breed.* 36 (1), 4–8. doi:10.1007/s12975-014-0384-4
- Chung, C. L., Longfellow, J. M., Walsh, E. K., Kerdieh, Z., Van Esbroeck, G., Balint-Kurti, P., et al. (2010b). Resistance loci affecting distinct stages of fungal pathogenesis: use of introgression lines for QTL mapping and characterization in the maize--*Setosphaeria turcica* pathosystem. *BMC Plant Biol.* 10, 103. doi:10.1186/1471-2229-10-103
- Cimmyt, (2005). *Laboratory protocols*. CIMMYT applied molecular genetics laboratory.
- 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 (9), 1018–1025. doi:10.1094/phyto.2000.90.9.1018
- Crossa, J., Pérez-Rodríguez, P., Cuevas, J., Montesinos-López, O., Jarquín, D., de los Campos, G., et al. (2017). Genomic selection in plant breeding: methods, models, and perspectives. *Trends Plant Sci.* 22 (11), 961–975. doi:10.1016/j.tplants.2017.08.011
- 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. Mycol.* 55, 189–197. doi:10.3114/sim.55.1.189
- Decreux, A., Thomas, A., Spies, B., Brasseur, R., Van Cutsem, P., and Messiaen, J. (2006). *In vitro* characterization of the homolacturonan-binding domain of the wall-associated kinase WAK1 using site-directed mutagenesis. *Phytochemistry* 67 (11), 1068–1079. doi:10.1016/j.phytochem.2006.03.009
- Dekkers, J. C. (2007). Prediction of response to marker-assisted and genomic selection using selection index theory. *J. Anim. Breed. Genet.* 124 (6), 331–341. doi:10.1111/j.1439-0388.2007.00701.x
- Ding, J., Ali, F., Chen, G., Li, H., Mahuku, G., Yang, N., et al. (2015). Genome-wide association mapping reveals novel sources of resistance to northern corn leaf blight in maize. *BMC Plant Biol.* 15, 206. doi:10.1186/s12870-015-0589-z
- Du, L., Zhang, H., Xin, W., Ma, K., Du, D., Yu, C., et al. (2021). Dissecting the genetic basis of flowering time and height related-traits using two doubled haploid populations in maize. *Plants (Basel)* 10 (8), 1585. doi:10.3390/plants10081585
- Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., et al. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* 6, e19379. doi:10.1371/journal.pone.0019379
- Freyermark, P. J., Lee, M., Woodman, W. L., and Martinson, C. A. (1993). Quantitative and qualitative trait loci affecting host-plant response to *Exserohilum turcicum* in maize (*Zea mays* L.). *Theor. Appl. Genet.* 87 (5), 537–544. doi:10.1007/bf00221876
- Galiano-Carneiro, A. L., Kessel, B., Presterl, T., and Miedaner, T. (2021). Intercontinental trials reveal stable QTL for Northern corn leaf blight resistance in Europe and in Brazil. *Theor. Appl. Genet.* 134 (1), 63–79. doi:10.1007/s00122-020-03682-1
- Galiano-Carneiro, A. L., and Miedaner, T. (2017). Genetics of resistance and pathogenicity in the maize/*Setosphaeria turcica* pathosystem and implications for breeding. *Front. Plant Sci.* 8, 1490. doi:10.3389/fpls.2017.01490
- Glaubitz, J. C., Casstevens, T. M., Lu, F., Harriman, J., Elshire, R. J., Sun, Q., et al. (2014). TASSEL-GBS: a high-capacity genotyping by sequencing analysis pipeline. *PLOS ONE* 9 (2), e90346. doi:10.1371/journal.pone.0090346
- Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., et al. (2012). Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* 40, D1178–D1186. doi:10.1093/nar/gkr944
- He, W., Yang, L., Leng, Y., Zhang, B., Yang, J., Li, L., et al. (2018). QTL mapping for resistance of maize to grey leaf spot. *J. Phytopathology* 166 (3), 167–176. doi:10.1111/jph.12673

- Human, M. P., Berger, D. K., and Crampton, B. G. (2020). Time-course RNAseq reveals *Exserohilum turcicum* effectors and pathogenicity determinants. *Front. Microbiol.* 11, 360. doi:10.3389/fmicb.2020.00360
- Hurni, S., Scheuermann, D., Krattinger, S. G., Kessel, B., Wicker, T., Herren, G., et al. (2015). The maize disease resistance gene Htn1 against northern corn leaf blight encodes a wall-associated receptor-like kinase. *Proc. Natl. Acad. Sci. U. S. A.* 112 (28), 8780–8785. doi:10.1073/pnas.1502522112
- Jaccoud, D., Peng, K., Feinstein, D., and Kilian, A. (2001). Diversity arrays: a solid state technology for sequence information independent genotyping. *Nucleic Acids Res.* 29 (4), E25. doi:10.1093/nar/29.4.e25
- Jamann, T. M., Balint-Kurti, P. J., and Holland, J. B. (2015). QTL mapping using high-throughput sequencing. *Methods Mol. Biol.* 1284, 257–285. doi:10.1007/978-1-4939-2444-8_13
- Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28 (1), 27–30. doi:10.1093/nar/28.1.27
- Khan, S. U., Saeed, S., Khan, M. H. U., Fan, C., Ahmar, S., Arriagada, O., et al. (2021). Advances and challenges for QTL analysis and GWAS in the plant-breeding of high-yielding: a focus on rapeseed. *Biomolecules* 11 (10), 1516. doi:10.3390/biom11101516
- Kibe, M., Nair, S. K., Das, B., Bright, J. M., Makumbi, D., Kinyua, J., et al. (2020a). 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, 572027. doi:10.3389/fpls.2020.572027
- Kibe, M., Nyaga, C., Nair, S. K., Beyene, Y., Das, B., Suresh, L. M., et al. (2020b). 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 (18), 6518. doi:10.3390/ijms21186518
- Kilian, A., Wenzl, P., Huttner, E., Carling, J., Xia, L., Blois, H., et al. (2012). Diversity arrays technology: a generic genome profiling technology on open platforms. *Methods Mol. Biol.* 888, 67–89. doi:10.1007/978-1-61779-870-2_5
- Kinyua, Z. M., Smith, J. J., Kibata, G. N., Simons, S. A., and Langat, B. C. (2010). Status of grey leaf spot disease in Kenyan maize production ecosystems. *Afr. Crop Sci. J.* 18 (4), 183–194. 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 Sci.* 25 (1), 192–194. doi:10.2135/cropsci1985.0011183x002500010046x
- Kolkman, J. M., Strable, J., Harline, K., Kroon, D. E., Wiesner-Hanks, T., Bradbury, P. J., et al. (2020). Maize introgression library provides evidence for the involvement of *liguleless1* in resistance to northern leaf blight. *G3 (Bethesda)* 10 (10), 3611–3622. doi:10.1534/g3.120.401500
- Koonin, E. V. (1993). *Escherichia coli* dinG gene encodes a putative DNA helicase related to a group of eukaryotic helicases including Rad3 protein. *Nucleic Acids Res.* 21 (6), 1497. doi:10.1093/nar/21.6.1497
- Korsman, J., Meisel, B., Kloppe, F. J., Crampton, B. G., and Berger, D. K. (2012). Quantitative phenotyping of grey leaf spot disease in maize using real-time PCR. *Eur. J. Plant Pathology* 133 (2), 461–471. doi:10.1007/s10658-011-9920-1
- Kotze, R. G., van der Merwe, C. F., Crampton, B. G., and Kritzing, Q. (2019). A histological assessment of the infection strategy of *Exserohilum turcicum* in maize. *Plant Pathol.* 68 (3), 504–512. doi:10.1111/ppa.12961
- Kump, K. L., Bradbury, P. J., Wissner, R. J., Buckler, E. S., Belcher, A. R., Oropeza-Rosas, M. A., et al. (2011). Genome-wide association study of quantitative resistance to southern leaf blight in the maize nested association mapping population. *Nat. Genet.* 43 (2), 163–168. doi:10.1038/ng.747
- Landi, P., Giuliani, S., Salvi, S., Ferri, M., Tuberosa, R., and Sanguineti, M. C. (2010). Characterization of root-yield-1.06, a major constitutive QTL for root and agronomic traits in maize across water regimes. *J. Exp. Bot.* 61 (13), 3553–3562. doi:10.1093/jxb/erq192
- 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 (5), 797–803. doi:10.1007/s001220100599
- Lehti-Shiu, M. D., and Shiu, S. H. (2012). Diversity, classification and function of the plant protein kinase superfamily. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 367 (1602), 2619–2639. doi:10.1098/rstb.2012.0003
- Lennon, J. R., Krakowsky, M., Goodman, M., Flint-Garcia, S., and Balint-Kurti, P. J. (2017). Identification of teosinte alleles for resistance to southern leaf blight in near isogenic maize lines. *Crop Sci.* 57 (4), 1973–1983. doi:10.2135/cropsci2016.12.0979
- Leonard, K. J., and Suggs, E. G. (1974). *Setosphaeria proleta*, the Ascigerous state of *Exserohilum prolatum*. *Mycologia* 66 (2), 281–297. doi:10.1080/00275514.1974.12019603
- Lipka, A. E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P. J., et al. (2012). GAPIT: genome association and prediction integrated tool. *Bioinformatics* 28 (18), 2397–2399. doi:10.1093/bioinformatics/bts444
- Littell, R. C., Milliken, G., Stroup, W. W., Wolfinger, R., and Schabenberger, O. (2007). *SAS for mixed models*, 61. Cary, NC, USA: SAS Institute.
- 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 (2), 304–312. doi:10.1094/pdis-08-14-0825-re
- Lopez-Zuniga, L. O., Wolters, P., Davis, S., Weldekidan, T., Kolkman, J. M., Nelson, R., et al. (2019). Using maize chromosome segment substitution line populations for the identification of loci associated with multiple disease resistance. *G3 (Bethesda)* 9 (1), 189–201. doi:10.1534/g3.118.200866
- Lorenz, A. J., and Smith, K. P. (2015). Adding genetically distant individuals to training populations reduces genomic prediction accuracy in barley. *Crop Sci.* 55 (6), 2657–2667. doi:10.2135/cropsci2014.12.0827
- 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., Wissner, R. J., Holland, J. B., and Balint-Kurti, P. (2019). Validation and characterization of maize multiple disease resistance QTL. *G3 (Bethesda)* 9 (9), 2905–2912. doi:10.1534/g3.119.400195
- McDonald, B. A., and Linde, C. (2002). Pathogen population genetics, evolutionary potential, and durable resistance. *Annu. Rev. Phytopathol.* 40, 349–379. doi:10.1146/annurev.phyto.40.120501.101443
- Meng, L., Li, H., Zhang, L., and Wang, J. (2015). QTL IciMapping: integrated software for genetic linkage map construction and quantitative trait locus mapping in biparental populations. *Crop J.* 3 (3), 269–283. doi:10.1016/j.cj.2015.01.001
- Miedaner, T., Boeven, A. L. G., Gaikpa, D. S., Kistner, M. B., and Grote, C. P. (2020). Genomics-assisted breeding for quantitative disease resistances in small-grain cereals and maize. *Int. J. Mol. Sci.* 21 (24), 9717. doi:10.3390/ijms21249717
- Munialo, S., Dahlin, S., Onyango, C., Kosura, W., Marstorp, H., and Oboro, I. (2019). Soil and management-related factors contributing to maize yield gaps in western Kenya. *Food Energy Secur.* 9. doi:10.1002/fes3.189
- Murithi, A., Olsen, M. S., Kwemoi, D. B., Veronica, O., 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.* 12, 767883. doi:10.3389/fgene.2021.767883
- Ng, A., and Xavier, R. J. (2011). Leucine-rich repeat (LRR) proteins: integrators of pattern recognition and signaling in immunity. *Autophagy* 7 (9), 1082–1084. doi:10.4161/auto.7.9.16464
- Nsibo, D. L., Barnes, I., Kunene, N. T., and Berger, D. K. (2019). Influence of farming practices on the population genetics of the maize pathogen *Cercospora zeina* in South Africa. *Fungal Genet. Biol.* 125, 36–44. doi:10.1016/j.fgb.2019.01.005
- Nsibo, D. L., Barnes, I., Omondi, D. O., Dida, M. M., and Berger, D. K. (2021). Population genetic structure and migration patterns of the maize pathogenic fungus, *Cercospora zeina* in East and Southern Africa. *Fungal Genet. Biol.* 149, 103527. doi:10.1016/j.fgb.2021.103527
- Nyanapah, J. O., Ayiecho, P. O., Nyabundi, J. O., Otieno, W., and Ojiambo, P. S. (2020). Field characterization of partial resistance to gray leaf spot in elite maize germplasm. *Phytopathology* 110 (10), 1668–1679. doi:10.1094/phyto-12-19-0446-r
- Peng, Y. L., Shirano, Y., Ohta, H., Hibino, T., Tanaka, K., and Shibata, D. (1994). A novel lipoxigenase from rice. Primary structure and specific expression upon incompatible infection with rice blast fungus. *J. Biol. Chem.* 269 (5), 3755–3761. doi:10.1016/s0021-9258(17)41924-7
- 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 (17), 6893–6898. doi:10.1073/pnas.1010894108
- 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 (6), 1729–1752. doi:10.1007/s00122-021-03773-7
- 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 (8), 904–909. doi:10.1038/ng1847
- Qi, H., Zhu, X., Shen, W., and Zhang, Z. (2023). A novel wall-associated kinase TaWAK-5D600 positively participates in defense against sharp eyespot and fusarium crown rot in wheat. *Int. J. Mol. Sci.* 24 (5), 5060. doi:10.3390/ijms24055060
- 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 (1), 21949. doi:10.1038/s41598-020-78928-5
- Ribaut, J. M., Hoisington, D. A., Deutsch, J. A., Jiang, C., and Gonzalez-de-Leon, D. (1996). Identification of quantitative trait loci under drought conditions in tropical maize. 1. Flowering parameters and the anthesis-silking interval. *Theor. Appl. Genet.* 92 (7), 905–914. doi:10.1007/bf00221905
- Saito, B. C., Silva, L., Andrade, J., and Goodman, M. (2018). Adaptability and stability of corn inbred lines regarding resistance to gray leaf spot and northern leaf blight. *Crop Breed. Appl. Biotech* 18, 148–154. doi:10.1590/1984-70332018v18n2a21

- Sanchez, D. L., Liu, S., Ibrahim, R., Blanco, M., and Lübberstedt, T. (2018). Genome-wide association studies of doubled haploid exotic introgression lines for root system architecture traits in maize (*Zea mays* L.). *Plant Sci.* 268, 30–38. doi:10.1016/j.plantsci.2017.12.004
- Sánchez-Sevilla, J. F., Horvath, A., Botella, M. A., Gaston, A., Foltá, K., Kilian, A., et al. (2015). Diversity arrays technology (DArT) marker platforms for diversity analysis and linkage mapping in a complex crop, the octoploid cultivated strawberry (*Fragaria × ananassa*). *PLoS One* 10 (12), e0144960. doi:10.1371/journal.pone.0144960
- Sansaloni, C., Franco, J., Santos, B., Percival-Alwyn, L., Singh, S., Petroli, C., et al. (2020). Diversity analysis of 80,000 wheat accessions reveals consequences and opportunities of selection footprints. *Nat. Commun.* 11 (1), 4572. doi:10.1038/s41467-020-18404-w
- Shaner, G. E. (1977). The effect of nitrogen fertilization on the expression of slow-mildewing resistance in knox wheat. *Phytopathology* 77, 1051. doi:10.1094/phyto-67-1051
- Shi, L.-y., Li, X.-h., Hao, Z.-f., Xie, C.-x., Ji, H.-l., Lü, X.-l., et al. (2007). Comparative QTL mapping of resistance to gray leaf spot in maize based on bioinformatics. *Agric. Sci. China* 6 (12), 1411–1419. doi:10.1016/S1671-2927(08)60002-4
- 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 sorghii* Schwein). *Crop Sci.* 60 (6), 2971–2989. doi:10.1002/csc2.20345
- Sun, H., Zhai, L., Teng, F., Li, Z., and Zhang, Z. (2021). qRgls1.06, a major QTL conferring resistance to gray leaf spot disease in maize. *Crop J.* 9, 342–350. doi:10.1016/j.cj.2020.08.001
- Swart, V., Crampton, B. G., Ridenour, J. B., Bluhm, B. H., Olivier, N. A., Meyer, J. M., et al. (2017). Complementation of CTB7 in the maize pathogen *Cercospora zeina* overcomes the lack of *in vitro* cercosporin production. *Mol. Plant Microbe Interact.* 30, 710–724. doi:10.1094/MPMI-03-17-0054-R
- Tang, Y., Liu, X., Wang, J., Li, M., Wang, Q., Tian, F., et al. (2016). Gapit version 2: an enhanced integrated tool for genomic association and prediction. *Plant Genome* 9, 2. doi:10.3835/plantgenome2015.11.0120
- 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 (Bethesda)* 3 (2), 197–203. doi:10.1534/g3.112.004630
- Tuberosa, R., Salvi, S., Sanguineti, M. C., Landi, P., Maccaferri, M., and Conti, S. (2002). Mapping QTLs regulating morpho-physiological traits and yield: case studies, shortcomings and perspectives in drought-stressed maize. *Ann. Bot.* 89 (7), 941–963. doi:10.1093/aob/mcf134
- Van Inghelandt, D., Melchinger, A. E., Martinant, J. P., and Stich, B. (2012). Genome-wide association mapping of flowering time and northern corn leaf blight (*Setosphaeria turcica*) resistance in a vast commercial maize germplasm set. *BMC Plant Biol.* 12, 56. doi:10.1186/1471-2229-12-56
- Vivek, B., Kasango, J., Chisoro, S., and Magorokosho, C. (2007). *Fieldbook: software for managing a maize breeding program: a cookbook for handling field experiments, data, stocks and pedigree information*. Mexico, DF: CIMMYT.
- Wallace, J. G., Bradbury, P. J., Zhang, N., Gibon, Y., Stitt, M., and Buckler, E. S. (2014). Association mapping across numerous traits reveals patterns of functional variation in maize. *PLoS Genet.* 10 (12), e1004845. doi:10.1371/journal.pgen.1004845
- Wang, N., Yuan, Y., Wang, H., Yu, D., Liu, Y., Zhang, A., et al. (2020). Applications of genotyping-by-sequencing (GBS) in maize genetics and breeding. *Sci. Rep.* 10, 16308. doi:10.1038/s41598-020-73321-8
- Welz, G., and Geiger, H. (2000). Genes for resistance to northern corn leaf blight in diverse maize populations. *Plant Breed.* 119, 1–14. doi:10.1046/j.1439-0523.2000.00462.x
- Welz, H. G. (1998). *Genetics and epidemiology of the pathosystem Zea mays/setosphaeria turcica*.
- Wenzl, P., Carling, J., Kudrna, D., Jaccoud, D., Huttner, E., Kleinbols, A., et al. (2004). Diversity arrays technology (DArT) for whole-genome profiling of barley. *Proc. Natl. Acad. Sci. U. S. A.* 101 (26), 9915–9920. doi:10.1073/pnas.0401076101
- Wisser, R. J., Balint-Kurti, P. J., and Nelson, R. J. (2006). The genetic architecture of disease resistance in maize: a synthesis of published studies. *Phytopathology* 96 (2), 120–129. doi:10.1094/phyto-96-0120
- Wisser, R. J., Kolkman, J. M., Patzoldt, M. E., Holland, J. B., Yu, J., Krakowsky, M., et al. (2011). Multivariate analysis of maize disease resistances suggests a pleiotropic genetic basis and implicates a GST gene. *Proc. Natl. Acad. Sci. U. S. A.* 108 (18), 7339–7344. doi:10.1073/pnas.1011739108
- Xiao, Y., Liu, H., Wu, L., Warburton, M., and Yan, J. (2017). Genome-wide association studies in maize: praise and stargaze. *Mol. Plant* 10 (3), 359–374. doi:10.1016/j.molp.2016.12.008
- Zhang, H., Jing, W., Zheng, J., Jin, Y., Wu, D., Cao, C., et al. (2020). The ATP-binding cassette transporter OsPDR1 regulates plant growth and pathogen resistance by affecting jasmonates biosynthesis in rice. *Plant Sci.* 298, 110582. doi:10.1016/j.plantsci.2020.110582
- 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 (8), 1797–1808. doi:10.1007/s00122-012-1954-z
- 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 (4), 769–776. doi:10.1007/s00122-011-1745-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 (1), 72–79. doi:10.1094/phyto-100-1-0072



OPEN ACCESS

EDITED BY

Krishnanand P. Kulkarni,
Delaware State University, United States

REVIEWED BY

Bin Wang,
Shaoguan University, China
Kubilay Yıldırım,
Ondokuz Mayıs University, Türkiye
Vijay Sheri,
East Carolina University, United States
Abira Chaudhuri,
National Institute of Plant Genome
Research (NIPGR), India

*CORRESPONDENCE

A. K. Mall,
✉ ashutosh.mall@icar.gov.in

[†]These authors have contributed equally
to this work

RECEIVED 06 June 2023

ACCEPTED 18 October 2023

PUBLISHED 09 November 2023

CITATION

Misra V, Mall AK, Pandey H, Srivastava S
and Sharma A (2023), Advancements and
prospects of CRISPR/Cas9 technologies
for abiotic and biotic stresses in
sugar beet.

Front. Genet. 14:1235855.
doi: 10.3389/fgene.2023.1235855

COPYRIGHT

© 2023 Misra, Mall, Pandey, Srivastava
and Sharma. 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.

Advancements and prospects of CRISPR/Cas9 technologies for abiotic and biotic stresses in sugar beet

Varucha Misra^{1†}, A. K. Mall^{1*†}, Himanshu Pandey^{1,2†},
Santeshwari Srivastava¹ and Avinash Sharma³

¹ICAR-Indian Institute of Sugarcane Research, Lucknow, India, ²Khalsa College, Amritsar, India, ³Faculty of Agricultural Sciences, Arunachal University of Studies, Namsai, India

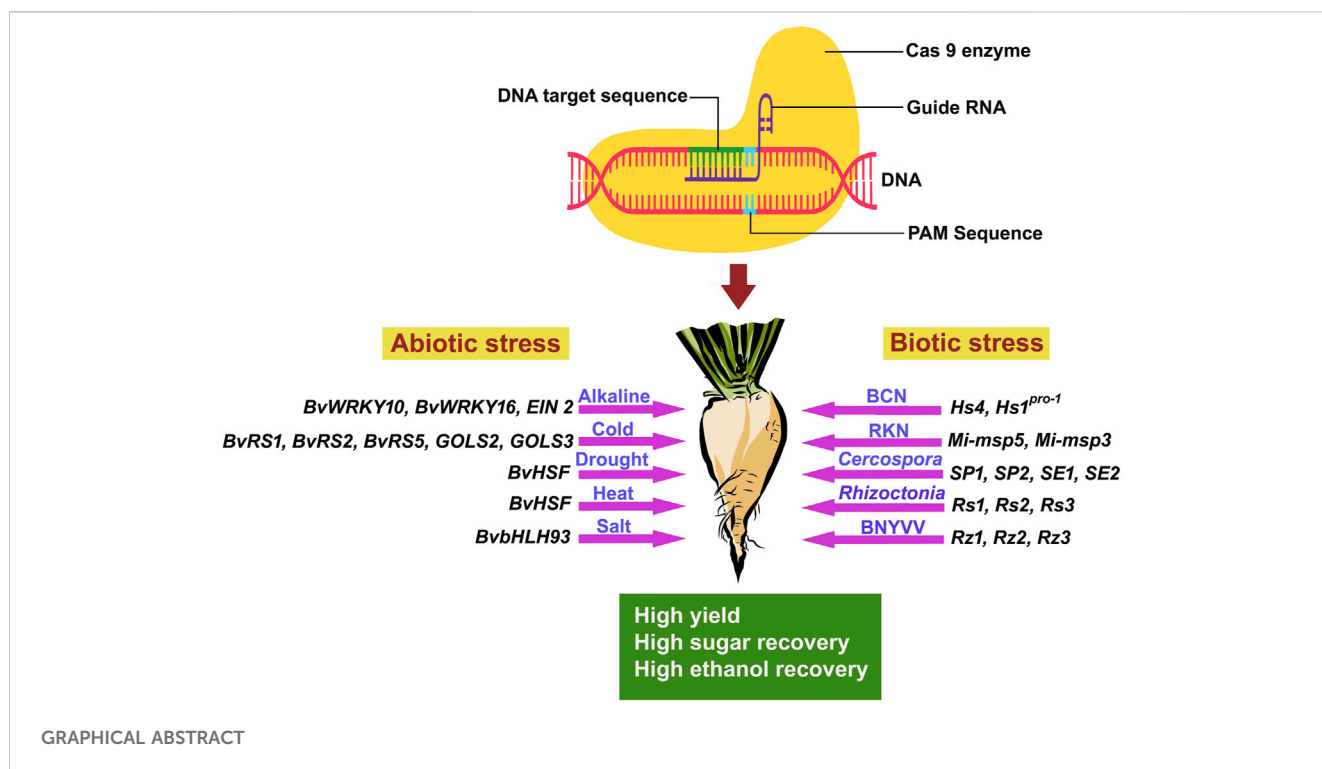
Sugar beet is a crop with high sucrose content, known for sugar production and recently being considered as an emerging raw material for bioethanol production. This crop is also utilized as cattle feed, mainly when animal green fodder is scarce. Bioethanol and hydrogen gas production from this crop is an essential source of clean energy. Environmental stresses (abiotic/biotic) severely affect the productivity of this crop. Over the past few decades, the molecular mechanisms of biotic and abiotic stress responses in sugar beet have been investigated using next-generation sequencing, gene editing/silencing, and over-expression approaches. This information can be efficiently utilized through CRISPR/Cas 9 technology to mitigate the effects of abiotic and biotic stresses in sugar beet cultivation. This review highlights the potential use of CRISPR/Cas 9 technology for abiotic and biotic stress management in sugar beet. Beet genes known to be involved in response to alkaline, cold, and heavy metal stresses can be precisely modified via CRISPR/Cas 9 technology for enhancing sugar beet's resilience to abiotic stresses with minimal off-target effects. Similarly, CRISPR/Cas 9 technology can help generate insect-resistant sugar beet varieties by targeting susceptibility-related genes, whereas incorporating *Cry1Ab* and *Cry1C* genes may provide defense against lepidopteron insects. Overall, CRISPR/Cas 9 technology may help enhance sugar beet's adaptability to challenging environments, ensuring sustainable, high-yield production.

KEYWORDS

abiotic, biotic, CRISPR/Cas, genes, genome editing, tolerance

1 Introduction

Sugar beet (*Beta vulgaris* L.) is cultivated in temperate regions of the world, however, its cultivation has spread to the tropical and subtropical zones of India (Mall et al., 2022a; Misra et al., 2022a). Sugar beet is well known for its sugar production in many countries of the world covering a significant area under cultivation (Table 1) and contributing to around 30% of the world's sugar requirement (Zicari et al., 2019). The root of sugar beet is an important source of natural sucrose as a sweetening agent (Mall et al., 2022b) and has diverse industrial applications (Misra et al., 2018). Sugar beet production faces major threats from biotic and abiotic stresses (Mulet, 2022). For instance, common viral (beet necrotic yellow vein virus, BNYVV (Ramchandran et al., 2021), bacterial (like *Pseudomonas aptata*), and fungal



diseases (like *Cercospora* (Misra et al., 2022c); *Alternaria* (Misra et al., 2020a), as well as nematodes (*Meloidogyne incognita*), and insects (like *Aphis fabae* (Mulet, 2022), *Spodoptera litura* (Santeshwari et al., 2020; Baitha et al., 2022)) hampers the production and productivity of the crop. Salinity, drought, heat (high temperatures), and cold are some of the abiotic stresses that severely impact sugar beet production all over the world (Misra et al., 2020b) (Table 2). Besides, other abiotic stresses like ozone build-up, flooding, nutritional deficiency, and heavy metal poisoning of the soil can also be challenging (Shabbir et al., 2022). In order for sugar beet plants to produce their optimum amount of sugar, enough moisture availability on a daily basis is necessary so as to allow for efficient transpiration and photosynthesis processes (Ober and Rajabi, 2010; Barratt et al., 2023). However, such ideal circumstances under natural environmental conditions have not been observed frequently owing to scanty rainfall or limited irrigating options. Despite the concerted efforts in developing tolerant varieties of sugar beet through conventional breeding and genome editing for improving the sugar and ethanol yield (Pattanayak et al., 2023), the scope still exists in the field of development of sugar beet varieties tolerance/resistance to abiotic and/or biotic stresses.

Molecular biology has witnessed a massive transformation due to the emergence and development of the CRISPR/Cas system as a biotechnological tool. The CRISPR/Cas 9 system is reportedly an efficient technology (Ahmad et al., 2021). The microbial adaptive immune system, CRISPR may target any genomic region by using a synthetic short guide RNA (sgRNA) (Jinek et al., 2012). Its strength comes in its capacity to effectively and precisely cause double-strand breaks in DNA at any location in the genome. CRISPR mediated genome editing can be used to change practically any sequence to expose its role in the genome (Asmamaw and Zawdie, 2021). On the

basis of the genes encoding the effector modules CRISPR–Cas systems can be classified where different cas proteins have unique characteristics and functional roles (Chaudhuri et al., 2022) (Table 3). CRISPR/Cas 9 mediated genome editing requires a protospacer adjacent motif (PAM) for Cas nuclease for initiation of the cutting process. The PAM is located 3–4 nucleotides downstream from the specific site where cleavage needs to be done (Gleditsch et al., 2019). The advances in genome editing techniques, particularly CRISPR/Cas 9 system, will benefit the cultivation of sugar beet by developing varieties resistant to abiotic and biotic stress conditions. This review discusses the current understanding of the mechanism of CRISPR/Cas 9 technology and its application in the improvement of sugar beet cultivars against abiotic/biotic stresses.

2 Site-directed nucleases (SDNs) and comparison of zinc finger nucleases, transcription activator like effector nucleases with CRISPR/Cas 9 technology

Targeted genome engineering has been emerged as an alternative to traditional plant breeding approaches, aiming to achieve a variety of crop improvement goals and sustainable food production (Misra et al., 2018). Genetic engineering (GE) techniques have been deployed to enhance the quality attributes of sugar beet (e.g., shelf life) (Monteiro et al., 2018) and improve its tolerance to biotic and abiotic stresses (Wan et al., 2021). In this context, the application of site-directed nucleases (SDNs) has evolved as a suitable GE technique for introducing desirable characteristics into plants (Aglawe et al., 2018). Zinc finger nucleases (ZFNs), transcription activator like

TABLE 1 Sugar beet area, yield, and production in major sugar beet producing countries of the world.

Country	Area (ha)	Yield (100 gm/ha)	Production (t)	References
China	229,300	342,386	7,850,900	FAO (2022)
France	491,880	855,116	34,365,390	
Germany	390,700	817,645	31,945,400	
Iran	91,803	560,650	5,146,924.8	
Italy	27,910	541,279	1,510,710	
Poland	250,570	609,564	15,273,850	
Russia	993,830	414,575	412,016,686	
Turkiye	288,940	631,620	18,250,000	
Ukraine	226,600	478,989	10,853,880	
United States	448,230	743,813	33,339,950	

TABLE 2 Impact of abiotic/biotic stresses on sugar beet yield.

Stresses		Reduction in sugar beet root yield (%)	References
Abiotic	Salt stress	49.3	Anagholi et al. (2018)
		10–50	Bybordi (2010)
	Cold	77 (in dry matter reduction)	Jalilian et al. (2017)
	Drought	27 (50% less irrigation); 21 (50% less irrigation)	Ghaffari et al. (2022)
		5 (Northern Europe)	Pidgeon et al. (2001)
		30 (Southern Russia)	
Disease resistance			
Biotic	Beet cyst nematode (<i>Heterodera schachtii</i> Schmidt)	25 to 50	Agrios (2005)
		21 (Italy)	Greco et al. (1993)
		60	Grujicic (1958); Cooke (1987)
		70	Pylypenko and Kalatur (2015)
	<i>Cercospora</i> leaf spot	30	Tan et al. (2023)
		40	Smith and Ruppel (1973); Esh and Taghian (2022)
		20–25 (India)	
	Beet curly top virus (BCTV)	30	AgResearch Magazine (2016)
	Beet necrotic yellow vein virus (BNYVV)	90	Johansson (1985); Casarini (1999)
Insect Pests			
	Armyworm (<i>Spodoptera</i> spp.)	>25 (Foliage damage)	DiFonzo et al. (2006)

(TAL) effector nucleases (TALENs), and Clustered regularly interspaced short palindromic repeats and CRISPR-associated (CRISPR/Cas) technology are instances of SDN methods. SDNs enable precise changes at predetermined places in a genome, avoiding any unintended random mutagenesis. They offer unparalleled control over targeted genome alterations, proving to be more cost-effective and efficient than conventional plant breeding and genetic engineering methods. SDNs have the potential to aid in crop improvement and enhance food security in many sugar beet

producing countries (Table 4). SDNs are categorised into SDN1, SDN2, and SDN3 based on the outcomes of genomic alterations and double strand break repair (Ghouri et al., 2023). SDN1 does not require template and causes gene disruptions *via* InDels (small insertions or deletions of bases). SDN2 uses a homologous template to repair or modify the gene at one or more locations. SDN3 requires the use of a whole gene as a template and results in gene substitution or foreign DNA insertion (Chen and Gao, 2020).

TABLE 3 Different types or classes of Cas proteins, emphasizing their unique characteristics and functional roles.

Protein	Class	Type	Sub type	Process	Functions
Cas 1	1, 2	I, II, III, IV, V, VI	A, B, C, D, E, F, U	Spacer acquisition	DNase, bind RNA; Cleavage at the specific site; Allow insertion of new spacers into CRISPR arrays
Cas 2	1, 2	I, II, III, IV, V, VI			Specific to U-rich regions; Homologous to mRNA interferase
Cas 3 (Signature)	1	I		Target interference	DNA helicase endonuclease; Cutting the target DNA at specific sites due to endonuclease domain
Cas 4	1, 2	I, II		Spacer acquisition and regulation	RecB-like nuclease homologous to RecB; Exonuclease activity and binds with RecBCD
Cas 5	1	I		crRNA expression and target binding	RAMP protein, crRNA biogenesis; (small subunit protein) Catalysation of crRNA processing; binds to a large subunit of Cas 8 in type I and Cas 10 in type III
Cas 6	1	I, III			RAMP protein, crRNA biogenesis; (nuclease activity) Repeat specific RNase involves in crRNA processing
Cas 7	1	I			RAMP protein, crRNA biogenesis; Co-transcriptional RNA cleavage during interference
Cas 8	1	I			Large protein with McrA/HNH-nuclease domain; Homologue of Cas10 protein; large subunit binding occurs with cas5 subunit
Cas 9 (Signature)	2	II	A, B, C	crRNA Target interference	Large multidomain protein with McrA-HNH nuclease domain; (Type II signature protein) crRNA dependent nuclease; Mediates RNA-guided DNA cleavage; widely used as DNA nucleases for inducing site-specific DNA breaks
Cas 10 (Signature)	1	III	A, B, C, D	crRNA expression and interference	HD nuclease domain, palm domain, Zn ribbon; Genetic manipulation
Cas 12a	2	V	A	Spacer acquisition	Cleaves the both complementary strands of the targeted DNA segment using a single RuvC nuclease domain
Cas 12b	2	V	B	Target interference	Dual-RNA-guided DNA nuclease and requires tracrRNA for further processing; higher efficiency for gene activation; wider target site for gene suppression; prefers T-rich protospacer adjacent motifs (PAMs)
Cas 12c	2	V	C	crRNA expression	Target binding rather than target degradation
Cas 12d (Signature)	2	V	D	Target interference	Catalyzes DNA cleavage with short complementary untranslated RNA
Cas 12e (Signature)	2	V	E	Target interference	Target DNA unwinding
Cas 13	2	VI	A, B, C, D	crRNA expression	Interference activity; Degradation of mRNAs; impart tolerance against plant viruses

The CRISPR/Cas system is a convenient replacement for ZFNs and TALENs in generating targeted genomic alterations (Ahmad et al., 2021). Both ZFNs and TALENs are utilized to mutate genomes at specific loci (Boti et al., 2023). However, these systems require two distinct DNA binding proteins flanking the region of interest, each having a C-terminal *FokI* nuclease module. Custom proteins are necessary for targeting DNA sequences. The process of designing and constructing custom proteins in both these technologies (Zinc finger motifs in ZFNs, while DNA binding domains obtained from TALE in TALENs, are required) is laborious and time-consuming. In CRISPR-mediated genome editing, PAM is located 3–4 nucleotides downstream from the specific site where cleavage will occur (Gleditsch et al., 2019). Therefore, accessibility to the protospacer adjacent motif (PAM) site, as the main determining factor for altering any functioning sequence, is easier (Akram et al., 2023). CRISPR technology depends on RNA-guided sequences, which can be designed effortlessly (Bajpai et al., 2023).

The target specificity in ZFNs and TALENs is more challenging as zinc fingers or TALEs identify short DNA sequences and require the combination of multiple modules to target the desired region, which may result in off-targets (Jyoti et al., 2023). The probability of off-targets is very low in CRISPR because it utilizes guide RNA to target the desired site, which can be easily programmed. Furthermore, the application of custom protein engineering in ZFNs and TALENs causes lesser adaptation and more laborious efforts (Kalaitzandonakes et al., 2022) while altering in sgRNA is easier for targeting the desired gene in CRISPR technology. The use of sgRNA in CRISPR technology is also beneficial in the delivery stage, making it a simple process compared to ZFNs and TALENs (Pankaj and Kumar, 2023). Due to the complexity involved in ZFNs and TALENs, including design and construction, target specificity, flexibility, and delivery, these methods are not preferred compared to CRISPR (Bhatia et al., 2023).

TABLE 4 Legislation of different sugar beet producing countries for genome editing crops.

Countries	Legislation	SDN 1	SDN 2	SDN 3	Year approved	Crops approved	References
United States	Classified SDN-1,2 genome-edited crops are equivalent to traditional breeds	Deregulated	Case by case		2018	Corn	Schmidt et al. (2020); Herrera et al. (2017); Chaturvedi (2004); Lombardo and Grando. (2020)
						Tomato	
					2017	Soyabean	
					2016	Mushroom	
					2017	Flax	
China	The Ministry of Agriculture announced preliminary recommendations for evaluating the safety of genome-edited plants that do not include exogenous DNA.	Under development			Not Applicable	No crops approved	Mallapaty (2022); Haque et al. (2018); Chaturvedi (2004); Wang et al. (2015a)
India	According to the memorandum, working with genome-edited plants must be done with extreme caution until exogenous inserted DNA is no longer present. The guidelines apply to SDN-1 and SDN-2 genome-edited plants	Under development			Not Applicable	No approved crops	Buchholzer and Frommer (2023); Lombardo and Grando. (2020); Chaturvedi (2004)
Russia	According to Resolution of 22 April 2019 no. 479, provision of funding for genome editing and defined transgene-free modified crops as equivalent to those produced through conventional breeding	New policies are expected			Not Applicable	No approved crops	Dobrovidova (2019)
UK	Initially, genome-edited crops will be free from GMO field trial rules. Field testing defines “qualified higher plants” as genome-edited plants that might have been developed using standard breeding procedures or may have occurred naturally. It is predicted that genome-edited plants will then be able to successfully enter field trials and acquire commercial approval without the need for case-by-case review	Case by case			Not Applicable	No approved crops	Stokstad (2021)
Brazil	Products obtained through site-directed random mutation involving the joining of non-homologous ends (SDN1 mutation) or site-directed homologous repair involving one or few nucleotides (SDN2 mutation) meet the criteria established in Normative Resolution No. 16 to be designated as non-GMO on a case-by-case basis. according to the resolution's provisions, site-directed transgene insertions (SDN3 mutation) are classified GM. If the product is labelled as GMO, the developer must meet all biosafety regulations and will be approved only after the CTNBio risk assessment. If the product is labelled non-GMO, it can be registered using the same methods as conventional items	Deregulated	Deregulated (If not transgenic)	NA		No approved crops	Schiemann et al. (2020); Chaturvedi (2004); Lombardo and Grando. (2020)

3 Application of CRISPR/Cas 9 technology for abiotic stress resistance in sugar beet

Sugar beet (*Beta vulgaris* L.) production is drastically affected by biotic and abiotic factors, which reduce the rate of photosynthesis, expansion of the canopy, development of the root system, and consequently, the accumulation of sucrose content in the plant (Misra et al., 2022a). Abiotic stresses, including temperature

fluctuations, water scarcity, salinity, metal toxicity, and UV radiation, are harmful to the sugar beet crop, and severely affect its yield across the world (Yu et al., 2020). These stressors greatly restrict the distribution of sugar beet crops, affect their developmental processes, and decrease sugar beet productivity (Ober and Rajabi, 2010). Improved understanding of multiple molecular mechanisms, like pathway signalling, activation of transcription factors, transcript modification (post-transcriptional modification), translation of processed transcript, and protein

TABLE 5 Identified genes in other crops for providing abiotic stress tolerance through CRISPR/Cas technology.

Abiotic stress	Genes targeted	Crops	References
Alkaline stress	<i>OsPPa6</i>	<i>Oryza sativa</i>	Wang et al. (2019)
Cold stress	<i>OsPIN5b</i> , <i>GS3</i> , <i>OsMYB30</i> , <i>OsAnn5</i> , <i>OsAnn3</i> , <i>OsPRP1</i>	<i>Oryza sativa</i>	Zeng et al. (2020), Shen et al. (2017), Nawaz et al. (2019)
	<i>AtWRKY34</i>	<i>Arabidopsis thaliana</i>	Zou et al. (2010)
	<i>SICBF1</i>	<i>Solanum lycopersicum</i>	Li et al. (2018)
	<i>VvWRKY24</i>	<i>Vitis vinifera</i>	Wang et al. (2014)
	<i>BcWRKY46</i>	<i>Brassica campestris</i>	Wang et al. (2011)
Heat stress	<i>OsPDS</i> , <i>OsHSA1</i> , <i>OsNAC006</i> , <i>OsNA C006</i> , <i>OsPyl14/6</i>	<i>Oryza sativa</i>	Nandy et al. (2019), Qiu et al. (2018), Wang et al. (2020), Miao et al. (2018)
	<i>AtWRKY25/26</i> , <i>AtWRKY33</i> , <i>AtWRKY39</i>	<i>Arabidopsis thaliana</i>	Li et al. (2011), Jiang et al. (2008), Park et al. (2005)
Drought stress	<i>AtOST2</i> , <i>AtAREB1</i> , <i>AtAVP1</i> , <i>AtmiR169a</i>	<i>Arabidopsis thaliana</i>	Osakabe et al. (2016), Roca Paixão et al. (2019), Park et al. (2017), Zhao et al. (2016)
	<i>OsERA1</i> , <i>OsSAPK2</i> , <i>OsSRL1</i> , <i>OsSRL2</i> , <i>OsDST</i> , <i>OsNAC14</i> , <i>OsPUB67</i>	<i>Oryza sativa</i>	Ogata et al. (2020), Lou et al. (2017), Liebe et al. (2020), Santosh Kumar et al. (2020)
	<i>BnaA6.RGA</i>	<i>Brassica napus</i>	Wu et al. (2020)
	<i>BdWRKY36</i>	<i>Brachypodium distachyon</i>	Sun et al. (2014)
	<i>BcWRKY46</i> , <i>BnaA6.RGA</i>	<i>Brassica campestris</i>	Wang et al. (2011), Wu et al. (2020)
	<i>GmMYB118</i>	<i>Soyabean</i>	Du et al. (2018)
	<i>GmMYB118</i>	<i>Chickpea</i>	Badhan et al. (2021)
Salt stress	<i>AtWRKY</i> , <i>AtWRKY4</i> , <i>AtACQOS</i>	<i>Arabidopsis thaliana</i>	Li et al. (2021a), Kim et al. (2021)
	<i>OsDST</i> , <i>OsSPL10</i> , <i>OsRAV2</i> , <i>OsBBS1</i> , <i>OsNAC45</i> , <i>OsAGO2</i> , <i>OsVDE</i> , <i>OsRR22</i> , <i>OsSAPK2</i> , <i>OsPQT3</i> , <i>OsPIL14</i> , <i>OsBGE3</i>	<i>Oryza sativa</i>	Santosh Kumar et al. (2020), Lan et al. (2019), Duan et al. (2016), Zeng et al. (2018), Zhang et al. (2020b), Yin et al. (2020), Wang et al. (2003), Zhang et al. (2019), Mo et al. (2020), Alfatih et al. (2020), Yin et al. (2017)
	<i>VpWRKY1</i> , <i>VpWRKY2</i> , <i>VpWRKY3</i>	<i>Vitis pseudoreticulata</i>	Li et al. (2010), Zhu et al. (2012)
Herbicide stress	<i>OsTB1</i> , <i>OsALS</i> , <i>OsACC</i>	<i>Oryza sativa</i>	Butt et al. (2018), Zhang et al. (2020b), Lyu et al. (2022)
Metal stress	<i>Atox1</i>	<i>Arabidopsis thaliana</i>	Baeg et al. (2021)
	<i>OsARM1</i> , <i>OsNramp5</i> , <i>OsLCT1</i> , <i>OsHAK1</i> , <i>OsPRX2</i>	<i>Oryza sativa</i>	Wang et al. (2017b), Tang et al. (2017), Lu et al. (2017), Nieves-Cordones et al. (2017), Mao et al. (2019)

modifications after the translation process, underlying stress responses of sugar beet crops at multiple levels, would be helpful in increasing the sugar beet production and sustainability through the application of CRISPR/Cas 9 (Yu et al., 2020; Misra et al., 2022a).

Abiotic stress tolerance is a complex trait that is mediated by multiple genes. The components of metabolic, regulatory, and signalling networks in plants interact and crosstalk extensively under abiotic stress conditions (Garg et al., 2014; Mickelbart et al., 2015). Several abiotic stress resistance-conferring genes have been identified in plants (Table 5) and introgressed into related crops through the application of biotechnological tools (Razaq et al., 2021).

Gene activation or repression occurs by targeting transcriptional activator or repressor complexes to particular sites within the gene

promoter region with catalytically inactivated Cas endonuclease. For this purpose, CRISPR-based approaches like CRISPRa and CRISPRi could be beneficial in characterizing genes for abiotic stress tolerance (Osakabe et al., 2016). CRISPR activation, abbreviated as CRISPRa, is a CRISPR variation in which a catalytically dead (d) Cas 9 is coupled with a transcriptional effector to control target gene expression. When the guide RNA and the effector arm reach the genomic location, the dCas9 is unable to produce a cut, and the effector instead triggers downstream gene expression. CRISPRa technique is used to increase gene expression by targeting the promoter region upstream of the transcription start site (TSS), whereas the CRISPR interference (CRISPRi) approach suppresses gene expression by targeting the promoter region downstream of the TSS. The application of CRISPRa/i requires precise identification of

the TSS location (Davis et al., 2018; Thomas et al., 2019). CRISPRi is a technology that uses dCas9's programmable binding capacity to inhibit gene expression by preventing or interfering with RNA polymerase binding, transcription factor binding, and transcriptional elongation. sgRNA specific to a gene sequence's upstream regulatory region (e.g., promoter) or transcription initiation site could direct dCas9 to bind and inhibit transcription initiation or elongation, effectively silencing gene expression. To inhibit target gene expression, dCas9 is coupled with the transcription repression domain of the Kruppel associated box (KRAB). dCas9 alone or in combination with KRAB is an effective tool for knocking off one or more genes (Piatek et al., 2015). CRISPRi and CRISPRa regulate gene expression by increasing or suppressing RNA polymerase, respectively. Thus, CRISPRi and CRISPRa are promising approaches for exploring and regulating stress-regulatory genes, as well as developing abiotic stress tolerant varieties. CRISPRi/a has been utilized successfully in plants to modify expression by a factor of 1000 (La Russa and Qi, 2015). CRISPRi and CRISPRa technologies can also be used in sugar beet under abiotic stress conditions. Although there is a lack of information available on this aspect, there are prospects to be worked on. The application of these technologies can aid in regulating gene expression in response to abiotic stress conditions (McCarty et al., 2020). These could be correlated with gene expression in sugar beet under drought, salinity, or temperature stress to provide tolerance. Altering these gene expressions could produce resilient sugar beet varieties with respect to abiotic stress conditions. Inhibiting or activating the targeted stress-responsive genes in sugar beet under abiotic stress conditions will help in understanding the regulatory mechanism. This, in turn, will assist in developing sugar beet varieties tolerant to adverse environmental conditions.

Furthermore, Cas 12a could also be explored for targeting the specific genes contributing to abiotic stress pathways, like drought, salinity, temperature, etc. By altering the genes related to specific abiotic stress conditions in sugar beet, the plant's potential to adapt and survive under adverse conditions could be increased. For instance, Chen X. et al. (2018) reported the application of Cas12a in targeting drought-responsive genes (positive regulation of gene) in *Arabidopsis*. Altering these targeted drought-responsive genes through Cas 12a improved the plant's drought tolerance potential. Such plants exhibited better water retention capacity, reduction in wilting, and higher survivability under drought conditions.

The development and evolution of sugar beet varieties tolerant to abiotic stress induced by climate change and global warming are imperative (Yolcu et al., 2021). Therefore, CRISPR/Cas 9 technology is necessary to create highly resistant sugar beet varieties against biotic and abiotic stresses. The CRISPR/Cas 9 technology involves specific steps for developing novel sugar beet varieties for different abiotic stress tolerances (Figure 1).

3.1 Alkaline stress tolerance

Aside from ionic toxicity and osmotic stress, high pH in the alkaline soils disrupts cell pH stability, destroys cell membrane integrity, and reduces root vitality and photosynthetic activity (Fang et al., 2021). Under the condition of high saline stress, beet

varieties that are cultivated along with wild types (naturally occurring/non-domesticated beet varieties) exhibited high antioxidant enzyme activities (Wang et al., 2017; Li B. et al., 2021). Wu et al. (2019) reported that with increasing sodium bicarbonate concentrations, Na⁺ concentrations were enhanced significantly in shoots and roots of sugar beet plants under alkaline conditions while a steady level of potassium ion concentrations was observed. Maintenance of K⁺ and Na⁺ homeostasis could be a key strategy for sugar beets adjusting to alkaline stress. Several genes (like WRKY, NAC, MYB, etc.) have been reported to be involved in conferring abiotic stress tolerance in plants (Chinnusamy et al., 2006; Hennig, 2012; Khadiza et al., 2017).

Wu et al. (2019) reported 58 WRKY genes, and among them, 9 genes were found to be responsible for alkaline stress responses (~15 mM–100 mM NaCHO₃) in both shoot and root parts. It evidently proved the increased expression of the *BvWRKY10* gene (in the terminal and lateral shoots) and *BvWRKY16* gene (in roots) under alkaline stress. WRKY proteins have been known to be associated with the response of plants to biotic and abiotic stress conditions (Jiang et al., 2017). The quantitative alterations and tissue-specific expressions of different *BvWRKY* genes have clearly shown its implications for alkaline stress tolerance in sugar beet. Research to alter the expression of *BvWRKY* genes using CRISPR/Cas 9 technology is required for the development of stress-resistant varieties. The *BvWRKY* family genes are specific and play crucial roles in wider aspects of sugar beet development processes like germination, root development, photosynthesis, etc., including response to alkaline stress conditions (Wu et al., 2019). Therefore, these genes can be engineered through CRISPR/Cas 9 technology to develop alkaline stress-resistant varieties in sugar beet (Table 6). Furthermore, long noncoding RNAs (lncRNAs) in response to alkaline stress had also been identified and characterized in sugar beet leaves. Besides, the interactions of candidate genes and miRNAs with the lncRNAs under stress conditions have also been reported in sugar beet (Zou et al., 2020). The use of CRISPR/Cas 9 technology on these identified/characterized lncRNAs will be helpful in evaluating their function in sugar beet and its expression can be modified to provide alkali stress tolerance in sugar beet. Additionally, genes belonging to the bHLH (basic helix–loop–helix) family in sugar beet involved in salt stress tolerance have also been well known. Wang Y. et al. (2021) reported the *BvbHLH93* gene as a salt-responsive gene that has been shown to confer tolerance under salt stress conditions in sugar beet. By increasing antioxidant activity and decreasing ROS generation, the *BvbHLH93* gene modulates salt stress tolerance in sugar beets. Furthermore, the *BvbHLH93* gene's ability to reduce *RbohD* and *RbohF* gene expression through modulating polyamine metabolism warrants additional investigation in sugar beet. CRISPR/Cas 9 technology will help in understanding the role and expression of these genes in salt stress conditions. This will aid in knowing the regulation of this gene in conferring tolerance to sugar beet for salt stress conditions.

3.2 Cold stress tolerance

Chilling temperature affects sugar beet cultivation, production, and economic yield (Moliterni et al., 2015; Bhattacharya, 2022).

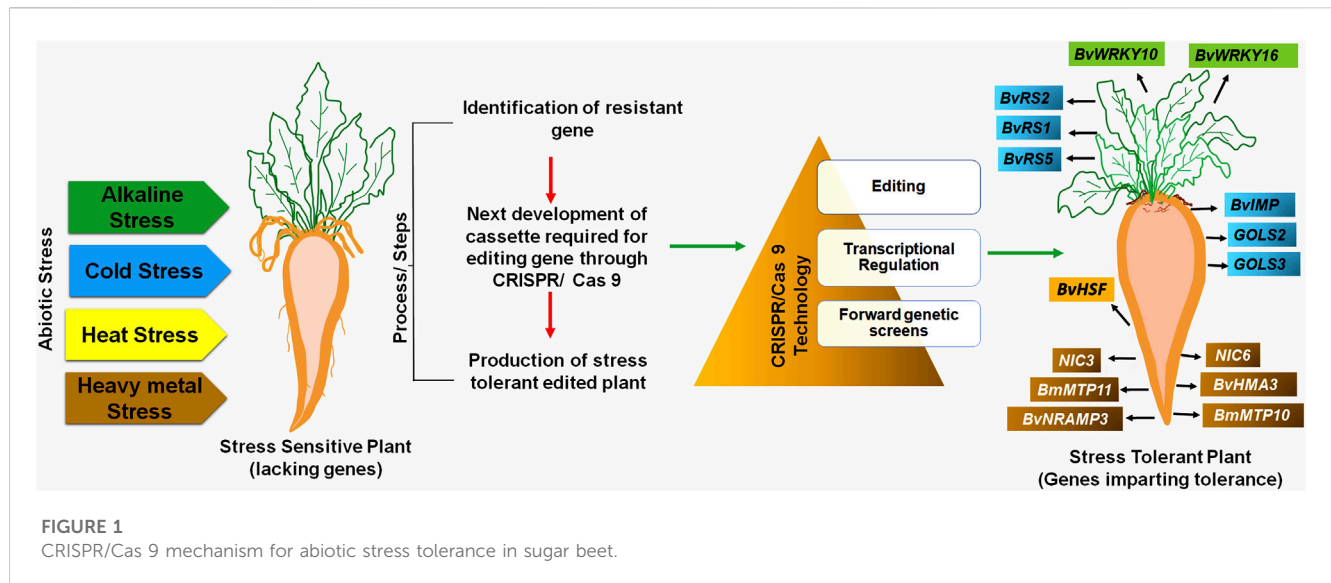


TABLE 6 Transcription factor that can be engineered through CRISPR/Cas 9 in sugar beet to provide resistance to different abiotic stress conditions.

Abiotic stresses	Prospective transcription factor that can be targeted for CRISPR/ Cas 9	Gene function	Gene expression	Editing mechanism	References
Alkaline	<i>BvWRKY10</i>	Provides resistance to alkaline and saline stress	Upregulation	Prime and base editing	Wu et al. (2019)
	<i>BvWRKY16</i>				Li et al. (2020a)
	<i>EIN 2</i>	• Ethylene insensitive protein	Induce/Increase	Prime and base editing	Lei et al. (2020)
		• Regulates ethylene response			Zhang et al. (2008)
		• Provides tolerance even against green peach aphid			Zhou et al. (2020)
Salt	<i>BvbHLH93</i>	• Provides resistance to salt stress	High expression	Prime and base editing	Wang et al. (2021a); Wang et al. (2021b)
		• Modulates salt stress tolerance			
		• Reduces <i>RbohD</i> and <i>RbohF</i> gene expression			
Chilling/Cold stress	<i>BvRS1</i>	• Stress genes responsible for raffinose synthase	Induce	Prime and base editing	Kito et al. (2018)
		• Provide resistance			
		• Encodes protein (comprises 783 amino acids)			
	<i>GOLS2</i>	• Genes responsible for galactinol synthase	Upregulation	Prime and base editing	Keller et al. (2021)
		• Provide resistance			
	<i>GOLS3</i>	• Stimulated by cold stress	Upregulation	Prime and base editing	
• Provide resistance					
Heat	<i>BvHSF</i>	• Genes responsible for higher expression of heat shock proteins during elevated temperature	Upregulation	Prime and base editing	Ismail et al. (2020)
		• Provide resistance			

Cold stress conditions have a serious impact on the sugar beets at various developmental stages, including early germination, sugar metabolism, growth, and bolting in the roots (Hoffmann and Kluge-Severin, 2011). The seedling stage of sugar beet is particularly prone to low temperature stresses. Cold stress during this stage leads to severe degeneration and growth retardation in the root system, consequently decreasing its sugar content (Moliteri et al., 2015; Jalilian et al., 2017). Porcel et al. (2018) revealed that overexpression of *BvCOLD1* gene in sugar beet exhibits cold tolerance potential along with other abiotic stress tolerance and overcoming boron deficiency. Keller et al. (2021) found that improved freezing-tolerant sugar beet genotypes accumulated more raffinose in the pith portion, which is vulnerable tissue to freeze damage. This finding demonstrated that raffinose and its precursors protect sugar beets from freezing damage. Recognizing the importance of raffinose in providing cold tolerance to sugar beet plants, Kito et al. (2018) isolated and characterized two genes, *BvRS1* and *BvRS2*, from the sugar beet plant. These genes code for the expression of raffinose synthase, which is a crucial enzyme during raffinose biosynthesis. An increase in transcription levels of *BvRS1* and *BvRS2* genes was observed in response to chilling stress in both leaves and roots (Kito et al., 2018). Furthermore, the production and build-up of oligosaccharides belonging to the raffinose family are particularly important during cold hardiness. Galactinol synthase (Gols) is considered a key regulator of the synthesis of such oligosaccharides and their accumulation (Vinson et al., 2020). The expression levels of *GOLS2* and *GOLS3* genes, responsible for galactinol synthase, as well as *BvRS2* and *BvRS5*, were high during chilling temperatures (Keller et al., 2021). Interestingly, the product of the *BvRS5* gene product and raffinose content increased exceptionally during freezing temperature in the taproots of tolerant sugar beet varieties, GT2 and GT3. In comparison with other sugar beet germplasm, GT2 exhibited high expression of *GOLS* and *RS* genes and raffinose content in roots, indicating chilling resistance in GT2 (Keller et al., 2021; Yolcu et al., 2021). Membrane proteins also frequently recognize cold stress conditions in sugar beet. These proteins activate a Ca^{2+} signal in the cytosol. Ca^{2+} -binding proteins may act as a bridge between the Ca^{2+} signal and several downstream transcription factors (Iqbal et al., 2022). It was reported that *B. vulgaris* *Integral Membrane Protein* gene resembles *AtERDL6*, which was previously reported for cold tolerance (Qi et al., 1995). Freezing conditions may be responsible for the increased transcription rate of *BvIMP* gene and vacuolar carbohydrates trafficking in sugar beet leaves, crucial for chilling stress response and germination of seed (Yu et al., 2020; Reyer et al., 2021).

Utilizing CRISPR/Cas technology, scientists have successfully enhanced cold stress tolerance in plants by editing genes allied with cold stress and raffinose synthesis. In the case of *Vitis vinifera*, the *VaDof17d* gene plays a critical role in the cold-responsive pathway and the production of raffinose family oligosaccharides. This is evidenced by the enhanced expression of galactinol synthase (*Gols*) and raffinose synthase genes. Mutating the *Dof17d-ED* gene through CRISPR/Cas 9 technology resulted in reduced cold tolerance and reduced levels of raffinose family oligosaccharides during cold stress (Wang et al., 2021). Therefore, genes including *BvIMP*, *BvRS1*, and *BvRS2* (Table 6) can be engineered in sugar beet through CRISPR/Cas 9 technology to develop site-specific mutants. These mutants

could be instrumental in enhancing cold tolerance in sugar beet crops by boosting their expression.

3.3 Heat stress tolerance

Increased temperatures and water scarcity tend to drastically affect the water content in plants where excess transpiration decreases the rate of water intake and causes permanent wilting in the plant (Kumar et al., 2004). Sugar beet is greatly affected by altered climatic conditions and disturbances in weather (Abou-Elwafa et al., 2020). High temperatures hamper major metabolic activities such as germination of seed, seed viability and its vigor, etc., leading to a threat to the survival of crop plants (Stevanato et al., 2019). Critical physiochemical processes, including photosynthesis and photosystem (PSII) activity also greatly affected due to blockage in the electron transport chain under high temperature stress (Murakami et al., 2000; Moore et al., 2021). The recent expansion of sugar beet cultivation in tropic and sub-tropical regions has drawn attention to farming it during the summer season (Abou-Elwafa et al., 2020).

Recognizing the importance of sugar beet cultivation across the world during the summer season, Ismail et al. (2020) explored the role of the transcription factor *BvHSF* gene, which showed higher expression levels under heat stress. The sugar beet crop showed enhanced expression levels in response to water scarcity, heat stress, and drought stress. Heat shock transcription factors (HSFs) are pivotal transcription factors in plants, critical for their response to various abiotic stresses such as heat, cold, salt, and drought (Fan et al., 2021). These HSF family members act by binding to the reverse repeat region of heat shock elements (HSEs), facilitating the transcription of heat shock proteins (HSPs) and assisting in the plant's stress adaptation mechanisms (Guo et al., 2016; Jacob et al., 2017). Wu et al. (2022) demonstrated that CRISPR/Cas 9 knockout mutants, specifically targeting single copy *MpHSF* genes (*MpHSF1V* and *MpHSF1B* mutants), resulted in different indel editing sites, showcasing enhanced thermotolerance in plants. Therefore, *BvHSF* genes can be engineered in sugar beet through CRISPR/Cas 9 technology to enhance their expression in sugar beet crops for increased tolerance towards heat resistance by creating in-deles (Table 6).

3.4 Drought stress tolerance

Drought stress also negatively impacts sugar beet root growth and development during the early phases of growth. Furthermore, the introduction of drought stress later in the growing season reduces leaf area and the number of leaves, ultimately resulting in lowered photosynthetic efficiency (Abou-Elwafa et al., 2020).

Generally, an increase in the expression of multiple drought-responsive genes and transcription factors enhances the plant's ability to tolerate drought conditions (Fang and Xiong, 2015; Kumar et al., 2020; Santosh Kumar et al., 2020). Conversely, upregulation of drought-sensitive genes in plants heightens their vulnerability to drought stress due to imbalances in hormonal levels, reduced antioxidant activities, and heightened production of reactive oxygen species (ROS). CRISPR/Cas 9 based genome

editing provides a promising avenue for enhancing drought tolerance in plants. This technique involves targeting negative regulators or drought-sensitive genes, allowing scientists to modify specific genetic elements and create crops that are more resilient to water scarcity. By precisely altering these genes, drought-resistant varieties can potentially be developed, ensuring sustainable agriculture in the face of changing environmental conditions. WRKY transcription factors are pivotal regulators of plant growth, development, and responses to both biotic and abiotic stresses. Among these factors, *WRKY3* and *WRKY4* genes in plants play a significant role in orchestrating the defense mechanisms against drought stress (Li B. et al., 2021). For instance, genetic manipulation of the *OsWRKY5* transcription factor has revealed significant insights into drought tolerance in plants. *OsWRKY5*, a key regulator, was found to hinder the plant's ability to withstand drought. During the seedling and heading phases, *OsWRKY5* was primarily expressed in growing leaves, and its expression decreased under drought stress conditions. Researchers conducted experiments using genome-edited loss-of-function alleles, *oswrky5-2* and *oswrky5-3*, to enhance drought tolerance. These edited alleles resulted in increased drought resistance, as evidenced by improved plant growth even under water scarcity (Lim et al., 2022). Conversely, when *OsWRKY5* was overexpressed in the activation-tagged line *oswrky5-D*, plants exhibited greater susceptibility to drought stress. Overexpression of *OsWRKY5* led to heightened sensitivity to abscisic acid (ABA), a plant hormone involved in stress response, and encouraged ABA-dependent stomatal closure. By editing the *OsWRKY5* genome, researchers successfully enhanced the plant's ability to produce grains even under drought stress conditions. This breakthrough offers valuable insights into improving crop resilience against water shortage, a critical factor in agricultural sustainability. Another example is for obtaining drought tolerance through CRISPR technology is enhancing the expression of *AREB1*, a specific transcription factor. In contrast, plants with a knocked-out *AREB1* gene exhibit increased sensitivity to drought stress (Singh et al., 2016). Roca Paixão et al. (2019) demonstrated enhanced drought stress tolerance through the utilization of CRISPR/dCas9 fusion with a Histone Acetyl Transferase (*AtHAT*) gene in *Arabidopsis*. These genes and transcription factors could also be targeted in sugar beet crops for attaining drought stress tolerance.

3.5 Heavy metal stress tolerance

Exposure of plants to toxic heavy metals causes different metabolic and physiochemical changes that depend on the concentration of these metals in soil, plant species, varieties, and abiotic conditions (Jamla et al., 2021; Thakur et al., 2022). Toxic metals like Pb damage the vacuolar membrane of sugar beet roots (Trela et al., 2012; Beata et al., 2022). Pb is one of the most toxic metals for plant cells, and it has a negative effect on the growth of plants, photosynthesis, respiration, and electron transport chain (Sharma and Dubey, 2005). Cd stress in *B. vulgaris* caused retarded growth, chlorosis, and increased root/plant ratio along with a decline in the rate of respiration in root-tips and photosynthesis (Greger and Ögren, 1991; Larbi et al., 2002; Liu et al., 2022). Haque et al. (2021) observed that higher levels of Cd in

sugar beet plants cause growth retardation due to an insufficient amount of Fe, resulting in decreased photosynthetic activity, and oxidative stress occurring in cells. Cd-treated plants display sensitivity to oxidative stress, leading to an increase in levels of O_2^- and H_2O_2 in roots and shoots. Additionally, Haque et al. (2021) reported the antioxidant defense mechanism in sugar beet under a higher concentration of toxic metal and observed that Cd stress enhances the activity of catalase (CAT) enzyme in the shoots, whereas the activities of superoxide dismutase (SOD), ascorbate peroxidase (APX), and glutathione reductase (GR) do not increase either in roots or in shoots. Genes involved in heavy metal stress tolerance in sugar beets have been explored and proven to be important. There are two MTP genes, *BmMTP10* and *BmMTP11*, reported for metal-resistant proteins from wild species of sugar beet (*B. maritima*). The detoxification process of Ni was controlled by genes from wild sugar beet (*B. maritima*) named as toxic nickel concentration (NIC), i.e., *NIC3*, *NIC6*, and *NIC8* (Bozdog et al., 2014; Yolcu et al., 2021). It is estimated that all these genes are required for protection against Ni toxicity. In a similar study, under Cd toxicity, sugar beet roots showed higher expression of putative *BvHMA3* and *BvNRAMP3* genes, suggesting that these genes are involved in the Cd uptake process (Haque et al., 2021).

Heavy metal-associated proteins (HMPs) and natural resistance-associated macrophage proteins (Nramp) are vital for heavy metal transport and detoxification within plant cells (Singh et al., 2016; Li W. et al., 2020). In rice, essential transporter genes such as *OsLCT1* and *OsNramp5* have been identified as key players in the absorption of Cd by the roots (Chang et al., 2020). Through CRISPR/Cas 9 enabled gene-expression manipulation, significant strides have been made in reducing the levels of Cd and Pb in rice grains. Specifically, the knockout of *OsNRAMP1* using CRISPR/Cas9 technology, as demonstrated in studies by Chu et al. (2022) and Wang F. Z. et al. (2017), has led to a substantial decrease in Cd and Pb content. Therefore, *BvHMA3* and *BvNRAMP3* genes can also be engineered in sugar beet through CRISPR/Cas 9 technology to develop heavy metal resistant varieties in sugar beet by creating site-specific mutagenesis.

4 Application of CRISPR/Cas 9 technology for improving biotic stress resistance

4.1 Pathogen resistance mechanism

The CRISPR/Cas 9 adaptive immune system for viral, bacterial, fungal resistance operates in three steps (Figure 2), regardless of the various shapes: 1) adaptation, 2) expression and maturity, and 3) interference. Protospacers, unique short DNA snippets from the invasive pathogen, are recognized by the Cas protein and inserted into CRISPR repeats as new spacers during the adaptation process. The host can then establish immunological memory and be equipped to recognize the same invasive infections in the future concerning this new spacer, which also serves as a genetic record (Paul et al., 2021).

For some CRISPR/Cas 9 systems to acquire the protospacer, the target DNA must have a short PAM (of 3-5 nucleotides) (Bolotin et al., 2005; Deveau et al., 2008; Shah et al., 2013). The CRISPR array

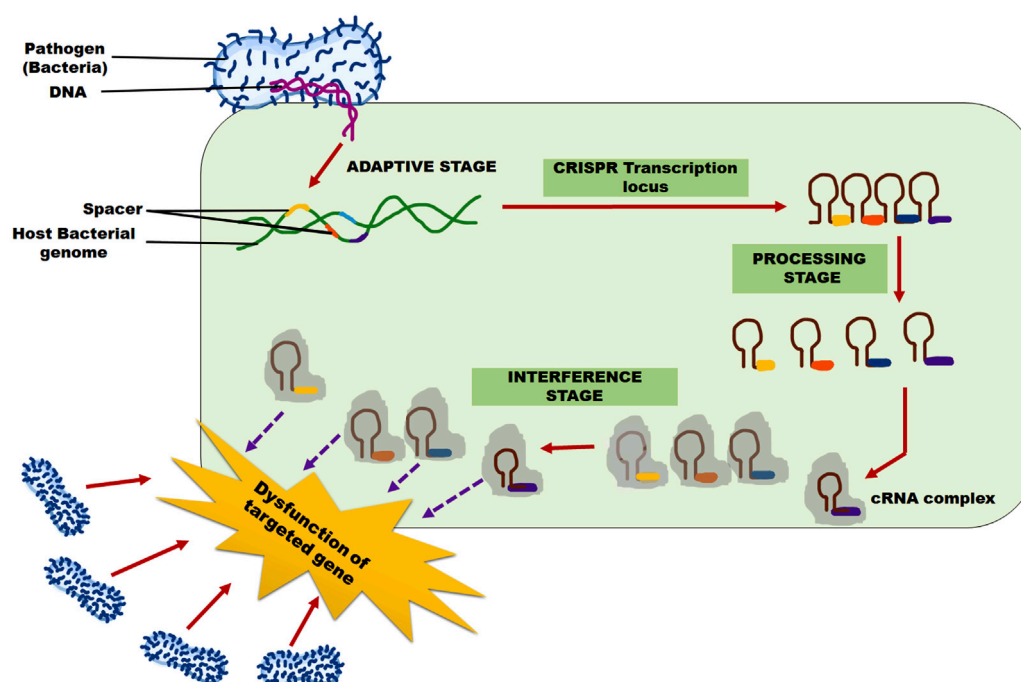


FIGURE 2

CRISPR/Cas 9 system mechanism for pathogen (Bacterial/Fungal/Viral) resistance in the plants. Adaptive, processing, and interference stages are three steps involved in this system.

is translated into a precursor-CRISPR RNA (pre-RNA), which is then processed to produce short mature CRISPR RNA (crRNA) through endo-nucleolytic cleavage and contains the sequences of the invading pathogen that have been learned (Carte et al., 2008; Haurwitz et al., 2010). At its 5' end, each crRNA has a single spacer (a brief RNA segment that complements the DNA sequence of the foreign genetic material), and at its 3' end, it has a CRISPR repeat sequence.

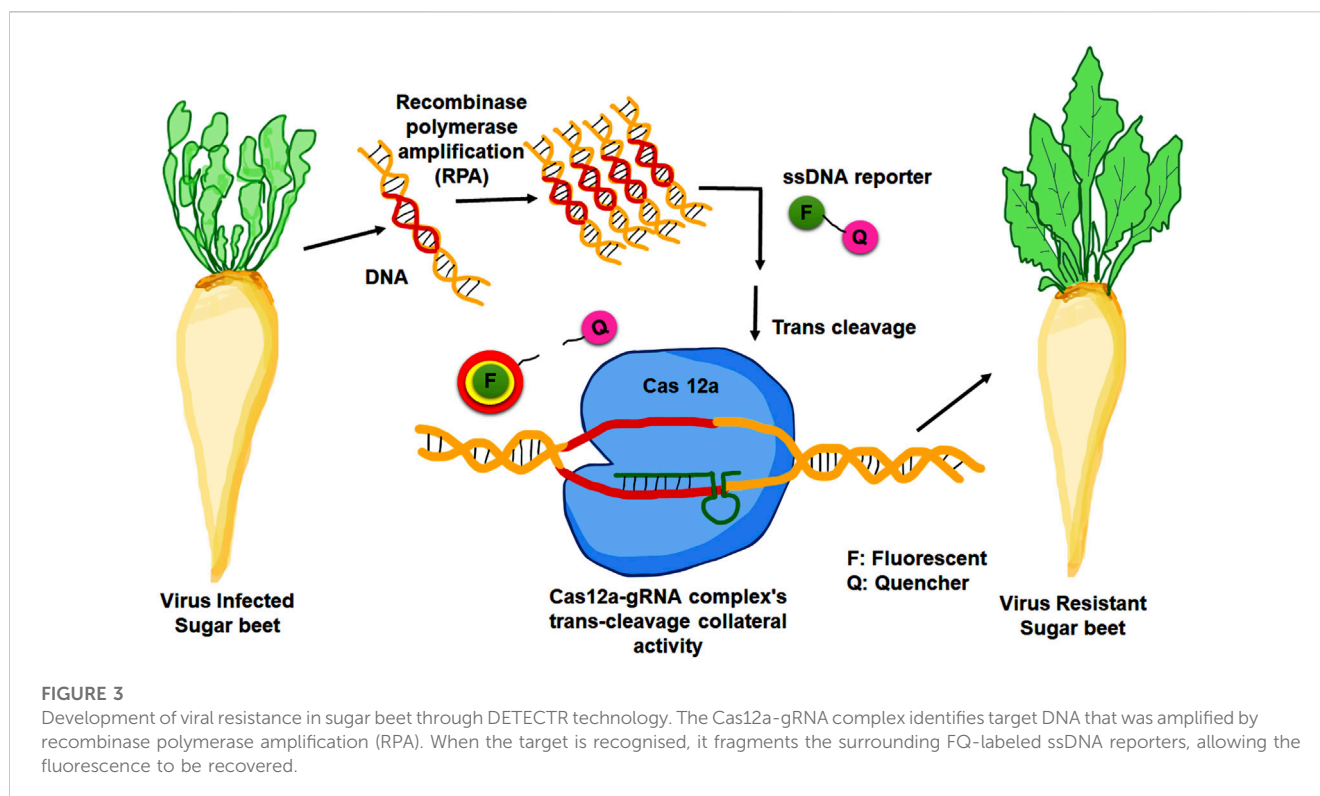
An active Cas-crRNA effector complex is formed when the mature crRNA combines with one or more Cas effector proteins. This complex searches for and attacks the cell's foreign nucleic acids during the interference phase. Using the specific PAM sequence either upstream or downstream of the protospacer, and relying on Watson-Crick base pairing, the crRNA component of the complex serves as a guide to identify the target DNA (Semenova et al., 2011; Jiang et al., 2013; Westra et al., 2013; Fineran et al., 2014; Zetsche et al., 2015). Once the target DNA is successfully recognized, the Cas nuclease cleaves and digests it. Different classes of the CRISPR immune system, based on various effector Cas proteins and PAM recognition sequences, have emerged in recent years.

Invading foreign DNA fragments of virus particles are recognized and eliminated by the CRISPR/Cas 9 system, allowing it to identify and remove DNA or RNA sequences that facilitate continued invasion (Barakate and Stephens, 2016). CRISPR/Cas 9 technology modifies the plant's inherent defense mechanism by detecting and removing harmful genes hidden within plant viruses. It can also be utilized to develop agricultural cultivars that are more resistant to specific plant viruses. This approach has fundamentally transformed virus resistance research because of its ability to use sequence-specific nucleases (Hsu et al., 2014).

4.2 CRISPR tools for disease diagnosis in sugar beet

The efficient implementation of control measures or management strategies heavily relies on the correct and timely identification of diseases and causative organisms. Consequently, disease diagnosis is critical and serves as the starting point for disease management. In this regard, Cas proteins play a significant role in managing diseases at the initial stages in plants. Cas proteins are attractive candidates for repurposing nucleic acid detection due to their programmability and extreme selectivity in binding and cleaving nucleic acids. Advances in understanding various Cas proteins have paved the way for the development of ultrasensitive, mobile, and cost-effective nucleic acid-based point-of-care (POC) testing equipment. Cas 9 proteins have been utilized to create robust and reliable nucleic acid detection technologies, but the recent discovery of Cas 13a and Cas 12a, with guaranteed cleavage activity, has revolutionized the field of nucleic acid detection (Gootenberg et al., 2017; Chen J. S. et al., 2018; Gootenberg et al., 2018; Azhar et al., 2021; Jiao et al., 2021).

A new molecular diagnostic approach known as DETECTR (DNA Endonuclease Targeted CRISPR Trans Reporter) technology (Chen J. S. et al., 2018), based on the CRISPR-Cas 12a system had been utilized for detecting BNYVV in sugar beet roots (Ramachandran et al., 2021). In this diagnostic method, Cas 12a cleaves any surrounding single-stranded DNA without regard for its target. This trait is known as collateral activity, and it has been exploited to develop DETECTR. In this approach, a complementary guide RNA first directs Cas12a to a target dsDNA (Figure 3). When Cas 12a binds to the correct target, it cleaves ssDNA reporter



molecules coupled with a quencher and a fluorophore. A fluorescence quencher (FQ)-labeled reporter was employed to monitor the trans-cleavage activity induced by the Cas12a-gRNA complex binding to the guide-complementary target DNA. In the presence of target DNA, the Cas 12a-gRNA complex's trans-cleavage activity is triggered, leading to the cleavage of surrounding FQ-labeled ssDNA reporters (Chen J. S. et al., 2018). The fluorescent signal generated by the separation of the quencher and the fluorophore detects the indiscriminate cleavage (Kocak and Gersbach, 2018). DETECTR exhibits enhanced sensitivity when combined with RPA preamplification. This diagnostic technique requires the amplification of viral fragments from template DNA under isothermal conditions. The one-step reverse transcriptase recombinase polymerase amplification (RPA) method was employed in sugar beet. The precise sensitivity of this method is its standout feature, as it can detect a single molecule of viral particle within a microliter of the sample (Kocak and Gersbach, 2018).

Another common CRISPR-based technology applied in sugar beet for viral resistance is nucleic acid sequence-based amplification CRISPR cleavage (NASBA). This technology helped the plants achieve total viral resistance. This technique involves targeting viral DNA using a guide RNA (gRNA) and cutting the viral DNA with the Cas 9 enzyme. The trans-activating crRNA (tracrRNA) base pairs with the repeat sequence in the crRNA to form a unique dual RNA hybrid structure guide that directs Cas9 to cleave the target DNA. A chimeric sgRNA combines crRNA and tracrRNA into a single RNA transcript. The two nuclease domains (RuvC and HNH) present in Cas 9 cut the target and non-target DNA strands, respectively. A short trinucleotide PAM is also required for the initial target sequence identification; without it,

the target sequence cannot be recognized. Successful identification results in a double-strand upstream of the 3'-NGG PAM (Li J. et al., 2022). This method is based on Cas 9 selective cleavage of target DNA and the toehold switch principle. This approach can distinguish between genotypes as it can identify a single base difference based on the presence or absence of a PAM sequence. This method utilizes sequence-based amplification, PAM-dependent target detection, Cas 9 cleavage, and a toehold sensor (Pardee et al., 2016).

Yildirim et al. (2019) utilized CRISPR/Cas 9 technology to confer various resistances in sugar beet against two curly top viruses (Beet curly top viruses (BCTV) and beet curly top Iranian viruses (BCTIV)). These viruses belong to two separate genera within the Geminiviridae family, specifically curtovirus and becurtovirus families, respectively. The gRNA/cas-9 endonuclease system was transiently overexpressed to check BCTV and BCTIV in sugar beet plants. Sugar beet plants overexpressing the gRNA/Cas 9 constructs exhibited decreased viral DNA accumulation, and this accumulation was also observed to be delayed compared to plants without the overexpression of gRNA/Cas 9 constructs. The CRISPR/Cas 9 system used in providing resistance to sugar beet plants against viral diseases, specifically targets the dsDNA of a geminivirus with gRNA. This approach helps restrict virus reproduction by disrupting critical replication genes (Khatodia et al., 2017). Ji et al. (2015) demonstrated the application of CRISPR technology against geminivirus in plants. Efficient antiviral sgRNAs can be identified to recognize specific sites in the viral genome. This approach has high potential for developing multiple resistances to all Geminiviruses. The viral resistance mechanism in sugar beet through the CRISPR/Cas 9 system is illustrated in the figure below (Figure 4).

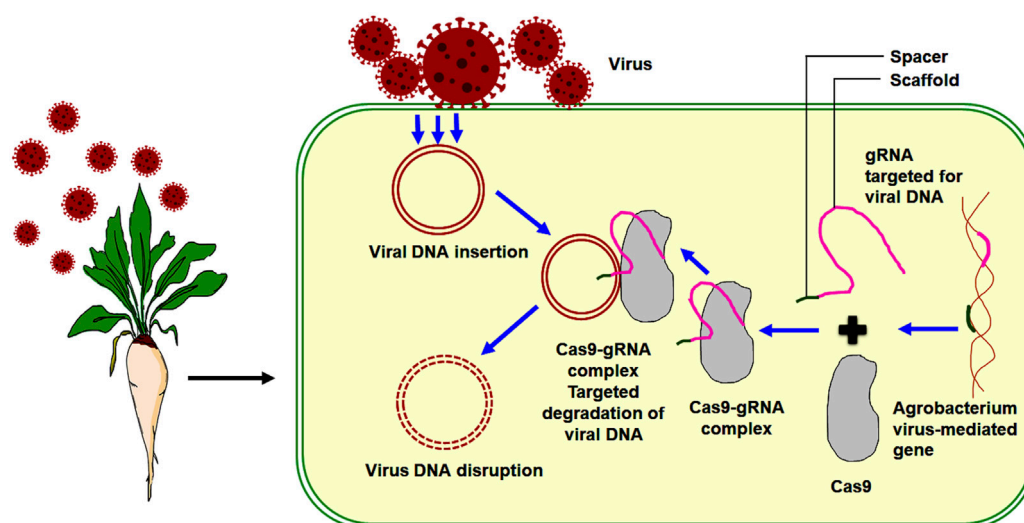


FIGURE 4

Viral resistance in sugar beet through CRISPR/Cas 9 system (Nucleic acid sequenced based amplification CRISPR cleavage).

The advent of CRISPR/Cas 9 has opened up numerous prospects for developing superior crop varieties through genome editing with great precision and accuracy. CRISPR techniques usher in a new era of breeding systems in which plant immunity is enhanced by disrupting the compatible connection between infections and hosts (Karmakar et al., 2022). CRISPR-based available tools like SHERLOCK, FLASH, and LEOPARD wherein Cas 12 and Cas 13 proteins have been utilized, will be helpful in the identification, diagnosis and management of diseases in sugar beet.

4.3 Beet cyst nematode (*Heterodera schachtii* schmidt) resistance

The host range of the beet cyst nematode (BCN) is wide and includes numerous species from abundant plant families, including Chenopodiaceae and Brassicaceae. In sugar beet cultivation (*Beta vulgaris* L.), *Heterodera schachtii* is a serious pest. It is known that cultivated *Beta* species lack the genes that provide resistance to nematodes. *Beta procumbens*, a wild species, and its allied species (*B. webbiana* and *B. patellaris*) are the sources of resistance genes (Cai et al., 1997). BCN is resistant to the *HsI^{pro-1}* locus. The native *HsI^{pro-1}* gene encodes a 282-amino acid protein with incomplete leucine-rich repeats and a potential membrane-spanning region. This protein is produced in sugar beet roots. The expression of the matching complementary DNA gave resistance to BCN infection in susceptible sugar beet (Cai et al., 1997). The *HsI^{pro-1}* (Table 7) promoter promotes nematode feeding site-specific GUS expression in both sugar beet and *Arabidopsis*, indicating a shared mechanism for regulating *HsI^{pro-1}* expression in these two species (Thurau et al., 2003).

Kumar et al. (2021) identified the resistance gene, *Hs4* (about 230 kb-sized area). CRISPR-Cas-mediated deletion (CRISPRi) and overexpression (CRISPRa) in susceptible sugar beet roots were used to characterize a candidate gene. The gene encodes a rhomboid-like protease predicted to be bound to the endoplasmic reticulum.

CRISPR Cas mutagenesis in the resistant sugar beet cultivar, NEMATA, deleted the ORF1. The roots of the knockout clones were extremely susceptible as huge numbers of J4 females and cysts packed with eggs observed in these roots. The expression of ORF in sugar beet roots determined nematode resistance/susceptibility. Beet roots expressing ORF 1 were resistant to nematodes while low ORF expression led to moderate susceptibility. Roots that did not express were susceptible to the nematode. Thus, CRISPRa can be a potential tool for the overexpressing targeted genes, which can confer tolerance to abiotic and biotic stress conditions (Horlbeck et al., 2016; Rai et al., 2019).

4.4 Beet necrotic yellow vein virus resistance

The multipartite genome of Beet necrotic yellow vein virus (BNYVV) consists of five positive-stranded RNAs, providing an enticing framework for the production of several foreign proteins (Jiang et al., 2019). Additionally, *NbPDS* guide RNAs were delivered through BNYVV-based vectors to transgenic plants expressing Cas9 for genome editing. This delivery resulted in a photobleached phenotype in systemically infected leaves. The BNYVV-based vectors will facilitate the expression and production of multiple proteins in sugar beet and related crop plants. gRNA can be delivered using BNYVV-based vectors for CRISPR/Cas 9 plant genome editing (Jiang et al., 2019).

In a breeding line developed by the Holly Sugar Company in the USA, partial resistance to BNYVV was found to be caused by a single dominant gene (*Rz1*) (Lewellen et al., 1987; Scholten et al., 1996). Accessions WB42 (*Rz2*) and WB41 (*Rz3*) of *Beta vulgaris* subsp. *maritima* from Denmark have also been identified as having BNYVV resistance (Lewellen et al., 1987; Gidner et al., 2005). Compared to the *Rz1* gene, the dominant *Rz2* gene seems to confer a higher level of resistance (Paul et al., 1993).

Detection of BNYVV is extremely sensitive and specific to the infected roots, as measured by the reporter signal. Ramachandran et al. (2021) established isothermal RT-RPA and CRISPR-based

TABLE 7 CRISPR/Cas 9 technology application in major sugar beet diseases for pathogen (bacterial/fungal/viral) resistance.

Diseases	Causal organism	Potential genes imparting tolerance/resistance	Gene function	Yield loss (%)	References
Nematode associated diseases					
Beet cyst nematode	<i>Heterodera schachtii</i>	<i>Hs1^{pro-1}</i>	Nematode resistance	Up to 60*	Cai et al. (1997)
		<i>Hs4</i>	Located on the wild beet translocation. Works together or independently of Hs1 ^{pro-1}		Kumar et al. (2021); *Ghaemi et al. (2020)
Root-knot nematode	<i>Meloidogyne incognita</i>	<i>R6m-1</i>	Nematode resistance	Up to 50%	Bakooie et al. (2015)
		<i>Mi-sbp-1</i>	Regulator of lipogenesis		Shivakumara et al. (2019)
		<i>Mi-cpl-1</i>	Interaction between plants and nematodes		Dutta et al. (2015)
		<i>Mi-msp3</i>	Nematode resistance		Joshi et al. (2020)
		<i>Mi-msp5</i>			
		<i>Mi-msp18</i>			
		<i>Mi-msp24</i>			
		<i>Hs1^{pro-1}</i>	General stress signalling genes and Nematode resistance		Abo-Ollo et al. (2018)
		<i>HSPRO2</i>			
		<i>Mi-1.2</i>	Nematode resistance		Bakooie et al. (2015)
		<i>R6m-1</i>			
Fungal diseases					
<i>Cercospora</i> leaf spot	<i>Cercospora beticola</i>	<i>SP1</i>	Acid chitinase activity	40*	Nielsen et al. (1994); Harverson, (2013)
		<i>SP2</i>			Nielsen et al. (1994)
		<i>SE1</i>	Chitinase activity		Nielsen et al. (1993)
		<i>SE2</i>	Exochitinase activity		Nielsen et al. (1993)
		<i>qcr1</i>	(Qualitative Trait Loci) QTL disease resistance		Taguchi et al. (2011)
		<i>qcr4</i>			
<i>Rhizoctonia</i> root rot	<i>Rhizoctonia</i> sp.	<i>Rs1</i>	QTL disease resistance	50*	Lein et al. (2007); *Barry (2006)
		<i>Rs2</i>			
		<i>Rs3</i>			
<i>Fusarium</i> root rot	<i>F. oxysporum</i>	<i>BvSP2</i>	SNP markers	40–50	Yerzhebayeva. (2018)
		<i>BvSE2</i>	Chitinase activity		
Powdery mildew	<i>Erysiphe betae</i>	<i>Pm 1</i>	Partial resistance	Up to 35*	Francis and Luterbacher (2003); Lewellen and Schrandt (2001); Grimmer et al. (2007); *Neher and Gallian (2013); *Francis (2002)
		<i>Pm 2</i>	Complete resistance to disease		
		<i>Pm 3</i>			
		<i>Pm 4</i>			
		<i>Pm 5</i>			
		<i>Pm 6</i>	Stronger resistance than others		
Aphanomyces seedling disease (Black rot or Black leg)	<i>Aphanomyces cochlioides</i>	<i>Acr 1</i>	Resistant gene	0–100	Taguchi et al. (2010); *Windels (2000)

(Continued on following page)

TABLE 7 (Continued) CRISPR/Cas 9 technology application in major sugar beet diseases for pathogen (bacterial/fungal/viral) resistance.

Diseases	Causal organism	Potential genes imparting tolerance/resistance	Gene function	Yield loss (%)	References
Viral diseases					
Viral Diseases	Virus Family	Vector	Potential Genes Imparting tolerance/resistance	Yield loss (%)	References
Beet severe curly top virus (BSCTV)	Geminiviridae	Leafhopper	<i>Rep</i>	100*	*Agrios (2005); Strausbaugh et al. (2017)
			<i>Ty</i>		Gupta et al. (2021)
Beet necrotic yellow vein virus (BNYVV)	Benyvirus	<i>Polymyxa betae</i>	<i>Rz1 (the Holly gene)</i>	90*	Lewellen et al. (1987); Paul et al. (1993); Liebe et al. (2020)
			<i>Rz2</i>		Scholten et al. (1999); Wetzel et al. (2021)
			<i>Rz3</i>		Francis and Luterbacher (2003); *Pferdmenges (2007)

virus diagnostic techniques for detecting BNYVV in sugar beet roots with rhizomania disease. Using this CRISPR-based technique, BNYVV in sugar beet roots baited for rhizomania can be identified, generating a readily identifiable fluorescence signal as compared to healthy reference root samples. The BNYVV RNA-1 sequence was chosen as the target since it is one of the least diverged components of BNYVV. The CRISPR-Cas 12a-based BNYVV detection approach has paved the way for a sensitive, focused, and high-throughput detection platform for the assessment of rhizomania (Ramachandran et al., 2021).

The creation and validation of such CRISPR-based BNYVV diagnostic approaches for sugar beet roots offer advantages in terms of sensitivity and resilience in isothermal circumstances. Hence, it would be a helpful tool for the companies involved in evaluating viruses to drive disease control strategies (Ramachandran et al., 2021). Furthermore, the technology established for virus identification in underground root tissue can be applied to create a CRISPR-based detection platform for viruses and soil-borne disease-causing pathogens in other crops as well.

4.5 Beet curly top virus resistance

Beet curly top virus is a single stranded DNA virus that belongs to the genus Curtovirus and has been known to cause outbreaks in sugar beet as a disease. This virus has a strong impact on sugar beet yield, reducing it to 30% and above. The beet leaf hopper (*Circulifer tenellus*) is the vector of this disease (Strausbaugh et al., 2012). Yildirim et al. (2019) utilized 20 gRNAs which targeted viral DNAs of this virus, and the plant was transformed with Cas 9 enzyme including vector (pKIR1.1) to provide antiviral resistance. Viral movement is inhibited by the gRNA/Cas 9 construct. Overexpressing the gRNA/cas-9 constructs in sugar beet plants resulted in a delayed and diminished accumulation of both viral DNA. In this manner, sugar beet was harnessed to build full viral resistance in the species.

Yildirim et al. (2022) used a genome-based characterization technology of beet curly top Iranian virus in sugar beets and

isolated and identified the Turkish variants of this virus along with its pathogenicity. Ebrahimi et al. (2022) explained the presence and pathogenicity of beet curly top Iranian virus in sugar beet. This helps in the management of this viral disease. Furthermore (Yildirim et al., 2023), reported that first-time broad-spectrum resistance against *Becurtovirus* using CRISPR technology where four gRNAs, involved in beet curly top Iranian virus, were cloned into vector consisting of Cas 9 and later agroinoculated into the virus infected leaves of sugar beet. Briddon et al. (1989) demonstrated that the genes for the capsid protein (CP), C4, and C2 are essential for viral pathogenesis and the emergence of disease symptoms in plants. The CP gene mutation caused the virus to lose its infectiousness and ability to propagate. Targeting these genes through the CRISPR/Cas 9 tool in sugar beet will help in the management of this viral disease.

The importance of miRNAs in sugar beet curly top virus resistance (BCTV) has been shown. Majumdar et al. (2022) determined that differentially expressed (DE) miRNAs are, in certain cases, only present in the R lines. Future functional evaluation of these potential miRNAs, either by overexpression in germplasm that is BCTV sensitive or by employing them as molecular markers to scan various sugar beet genomes, will aid in establishing BCTV resistance. Furthermore, Eini et al. (2022) demonstrated the development of very effective geminiviral replicons (GVR) from BCTV with a broad host range for recombinant gene expression in plants.

Prior to the advent of molecular markers, only a single disease resistance gene (C gene) for BCTV resistance had been mapped in sugar beet (Mutasa-Gottgens et al., 2000). Studies have successfully developed markers linked to resistance genes for BNYVV (Luterbacher et al., 1998; Luterbacher et al., 2000), BCN (Owen and Ryser, 1942; Scholten et al., 1997) powdery mildew (Uphoff and Wricke, 1992) and to quantitative trait loci against *Cercospora* leaf spot.

4.6 *Cercospora* leaf spot resistance

In response to infection by the leaf-spot fungus *Cercospora beticola*, two isoforms of acid chitinase (*SE1* and *SE2*) were found in the leaves of sugar beet. However, only one of the isoforms (*SE2*) had exochitinase

TABLE 8 Genes that can impart tolerance to sugar beet crop against insect pests (larval stage) via CRISPR/Cas 9 technology.

(a) Some genomes successfully edited in insects by the CRISPR/Cas tool kits						
Order	Insect name (common name)	Gene	Gene function	Editing	Outcome of the editing	References
		Imparting tolerance/resistance				
Lepidoptera	<i>Spodoptera litura</i> Fabricius (Armyworm)	<i>Slit PBP3</i>	Sex pheromone perception	Chimera mutation	Destroyed pest insect mating	Zhu et al. (2016)
		<i>siSe2</i> (Serine protease 2)	Sperm movement and activity	Knockout	Induces male sterility	Bi et al. (2022)
		<i>Abdominal-A (slabd-A)</i>	• Embryonic development gene	Knockout	Defected body segmentation and irregular pigmentation	Bi et al. (2016); Sun et al. (2017)
			• Body segmentation			
	<i>Spodoptera frugiperda</i> J.E. Smith (Fall armyworm)	<i>SlitBLOS2</i>	Acts as a marker gene	Knockout	Coloration of the integuments, a marker gene for functional studies and pest control strategies	Zhu et al. (2017)
		<i>BLOS2</i> (Biogenesis of lysosome-related organelles complex 1 subunit 2)		Knockout	Translucent mosaic integument	Zhu et al. (2020)
		<i>TO</i> (Tryptophan 2, 3-dioxygenase)			Olive eye color	
		<i>E93</i>	A key ecdysone-induced transcription factor that promotes adult development		Larval-pupal intermediate phenotypes	
		<i>ABCB1</i>	Susceptibility to chemical pesticides or Bt toxins	Knockout	Susceptibility to emamectinbenzoate, beta-cypermethrin and chlorantraniliprole	Li et al. (2022b)
	<i>Spodoptera exigua</i> (Beet armyworm)	<i>CYP9A186</i>	Restoration of Emamectin benzoate (EB) [(4''R)-4''-deoxy-4''-(methylamino) avermectin B1 benzoate] susceptibility	Knockout	Susceptibility to emamectin benzoate (EB)	Zuo et al. (2021)
		Ryanodine receptor	Regulates calcium release from intracellular stores and other cellular processes, viz., muscle contraction, gene transcription, neurotransmitter release, hormone secretion and cell proliferation	Substitution	Controlled insect population and resistance to various insecticides	Zuo et al. (2017)
		P-glycoprotein gene	Unknown	Knockout	Susceptibility to abamectin and emamectin benzoate	Jin et al. (2020)
		<i>Sea6</i> (<i>Spodoptera exigua</i> α-6-nicotinic acetylcholine receptor (nAChR))	Role in Spinosyn insecticide interaction	Knockout	Resistance to spinosyn insecticides	Zuo et al. (2018); Zuo et al. (2020)
	<i>Spodoptera littoralis</i> (Egyptian cotton leafworm)	Orco (Odorant receptor co-receptor)	Impairs feeding, mating and egg-laying behavior	Knockout	Reduced survival rate	Koutroumpa et al. (2016)
		<i>SlitOrco</i>	plant odor and sex pheromone olfactory detection	Knockout	Investigated the function of the <i>Orco</i>	Cui et al. (2017); Koutroumpa et al. (2016)
					gene in the non-model insect	
					<i>Spodoptera littoralis</i>	
	<i>Agrotis ipsilon</i> Rott. (Cutworm)	<i>Yellow-y</i>	Involved in body pigmentation and play a role in waterproofing	Knockout	Pigmentation plays a vital role in insect survival and reproduction	Chen et al. (2018a)
		<i>AiTH</i> (<i>Agrotis ipsilon</i> tyrosine hydroxylase)	Insect melanin and catecholamine biosynthesis pathway	Knockout	• Narrowing in the eggshell • Pigmentation of epidermis and newly hatched larval development	Yang et al. (2018)

(Continued on following page)

TABLE 8 (Continued) Genes that can impart tolerance to sugar beet crop against insect pests (larval stage) via CRISPR/Cas 9 technology.

(a) Some genomes successfully edited in insects by the CRISPR/Cas tool kits						
Order	Insect name (common name)	Gene	Gene function	Editing	Outcome of the editing	References
		Imparting tolerance/ resistance				
		<i>Aidsx</i>	Embryonic development	Knockout/ Disruption	Sexually dimorphic development and behavior	Chen et al. (2019)
		<i>AiMasc</i> (<i>Masculinizer gene</i>)	Sex determination		Induced expression of male specific double sex isoform	Wang et al. (2019)
	<i>Helicoverpa armigera</i> Hübner (Gram pod borer)	<i>GmUGT</i> (QTL-M)	Alteration of flavonoid biosynthesis pathway	Insertion	Resistance to insect	Zhang et al. (2022)
		HaCad (<i>Helicoverpa armigera</i> cadherin)	As receptor for Bt Cry 1A Toxin	Knockout	Resistance to Bt toxin Cry1Ac	Wang et al. (2016)
		Cluster of nine P450 genes	Defense mechanism against natural/synthetic insect toxins	Knockout	Identification of the key players in the insecticide metabolism	Wang et al. (2018)
		<i>CYP6AE</i>		Knockout	Regulation of detoxification enzymes	Wang et al. (2018)
		<i>OR16</i>	Pheromone antagonist	Knockout	Destroyed pest insect mating	Chang et al. (2017)
		<i>Tetraspanin</i>	Cell migration, signal transduction, and intracellular trafficking	Knockout	Resistance to Bt toxin cry1Ac	Jin et al. (2018)
		<i>HaABCA2</i>	Involved in the resistance mechanism for Cry2Ab	Knockout	Resistance to cry2Aa and cry2Ab	Wang et al. (2017a)
		<i>White</i>	Differential distribution of eye pigments	Knockout	Patterns of pigmentation	Khan et al. (2017)
		<i>Brown</i>			No phenotypic effects on viability or pigmentation	
		<i>Scarlet</i>			Increase amount of pteridines or ommochromes	
		<i>ok</i>	Analogous to that of brown gene in <i>Drosophila</i>	Knockout	Upregulation in gene expression during early larval instars	Zheng et al. (2020)
		<i>NPC1b</i>	Growth of <i>Helicoverpa armigera</i> larvae and dietary cholesterol uptake			

(b) Some genomes successfully edited in plants by the CRISPR/Cas tool kits against insects							
Order	Insect Name (Common Name)	Gene	Crop reported in	Gene Function	Editing	Outcome of the editing	References
		Imparting tolerance/ resistance					
Lepidoptera	<i>Spodoptera frugiperda</i> J.E. Smith (Fall armyworm)	<i>CryIFa</i>	Field crop	Practically resistance	Knockout	SeABCC2 has a major role and SeCad1 a minor role in mediating toxicity of Cry1Ac and Cry1Fa	Huanga et al. (2020) ; Flagel et al. (2018) ; Jin et al. (2020)
		<i>SfABCC2</i>		Resistance to <i>Cry1F</i> but no alteration in susceptibility to small molecule pesticides	Knockout	Cry1F due to mutations in the <i>SfABCC2</i> gene do not affect susceptibility to the synthetic and semisynthetic small molecule pesticides	Abdelgaffar et al. (2020) ; Jin et al. (2021)

(Continued on following page)

TABLE 8 (Continued) Genes that can impart tolerance to sugar beet crop against insect pests (larval stage) via CRISPR/Cas 9 technology.

(b) Some genomes successfully edited in plants by the CRISPR/Cas tool kits against insects							
Order	Insect Name (Common Name)	Gene	Crop reported in	Gene Function	Editing	Outcome of the editing	References
		Imparting tolerance/resistance					
		<i>Cry1Ac</i>	Field crop	(SfCad) cadherin protein	Knockout mutant strain	Targets different exons of the <i>SfCad</i> gene	Zhang et al. (2020a)
		<i>Cry1Ab</i>		Toxin protein			Zhang et al. (2020a); Kimoto and Shimamoto (2001); Hernández-Rodríguez et al. (2013); Jin et al. (2020)
		<i>Cry1C</i>		(SfCad) cadherin protein			
		<i>ABC transporters</i>		As receptor for Bt <i>Cry1Fa</i> and <i>Cry1Ab</i> toxins	Knock out	<ul style="list-style-type: none"> • Toxicity of two <i>Bacillus thuringiensis</i> <i>cry1</i> toxins to the pest 	Jin et al. (2020)
						Resistance to both <i>cry1fa</i> and <i>cry1ab</i> toxins	
	<i>Spodoptera littoralis</i> (Egyptian cotton leafworm)	<i>Cry1Ab</i>	Sugar beet	Toxin protein	Insertion	<ul style="list-style-type: none"> • Strong anti-feedant effect on insect 	Sedighi et al. (2011)
						<ul style="list-style-type: none"> • Increase in the developmental time and mortality 	
	<i>Helicoverpa armigera</i> Hübner (Gram pod borer)	<i>Cry1Ac</i>	Field crop	Cadherin as a receptor, Toxin protein	Insertion	Saturable, high-affinity binding to insect	Stewart et al. (2001); Wang et al. (2005); Wang et al. (2016); Chen et al. (2018b); Karim et al. (2000)
		<i>Cry2Ab</i>			Knockout	High levels of resistance	
		<i>GmUGT</i> (QTL-M)	Soyabean	Alteration of flavonoid biosynthesis pathway	Insertion	Overexpression of <i>GmUGT</i> produced sensitive soybean varieties against <i>H. armigera</i>	Zhang et al. (2022)
		<i>Cry2Aa</i>	Pigeon pea	Mediate toxicity	Knockout	High levels of resistance	Wang et al. (2017a); Singh et al. (2018)
Diptera	All species	<i>Cry2</i>	Multiple crops	Toxin protein	<ul style="list-style-type: none"> • Potential gene that can be targeted for CRISPR/Cas 9 • Transcription repression causes lack in ability to repress Clock: cycle mediated transcription 		Yuan et al. (2007)
	<i>Tetanops myopaeformis</i> (Sugar beet root maggot)	Delta-endotoxin genes	Sugar beet	Target an insect's digestive system	Insertion	Resistance	Wilhite et al. (2000); Smigocki et al. (2003)
Coleopteran	All species	<i>Cry3</i>	Multiple crops	Toxin protein	<ul style="list-style-type: none"> • Potential gene that can be targeted for insecticidal <i>Bt</i> toxins using CRISPR/Cas 9 		Ordaz et al. (1996); Pauchet et al. (2016)

activity and could successfully hydrolyze chito-oligosaccharides (Nielsen et al., 1993). In resistant vs. susceptible cultivars, the sugar beet *SE2* gene (*B. vulgaris* *SE2*) is expressed at a substantially higher level following *Cercospora* infection (Nielsen et al., 1993). The leaves of sugar beet infected by *C. beticola* also contained two isoforms of another acid chitinase (*SP1* and *SP2*), which are similar to *SE1* and *SE2*. Infected

sugar beet plant leaves had very high levels of *BvSP2* (*B. vulgaris* *SP2*) gene expression, however, *BvSP2* protein accumulation was only found in the vicinity of the infection sites (Nielsen et al., 1994).

Sugar beet leaves responded vigorously when exposed to a cell-wall protein solution made from the non-pathogenic oomycete *Pythium oligandrum* isolate. This response involved a substantial increase in

BvSE2 gene expression, which peaked at 4 h following the vaccination (Takenaka and Tamagake, 2009). However, since oomycetes lack chitin, *BvSE2* may be produced as part of a coordinated response to other proteins involved in the disease (Collinge et al., 1993).

4.7 Insect-pest resistance in sugar beet

Generally, resistance in insect-pest populations occur due to mutations in genes, which facilitates the encoding for receptor molecules and disrupt the interaction between the insect and the toxin. In rice, CRISPR/Cas9-dependent knockout of *CYP71A1* mutant gene enables it encode a functional tryptamine 5-hydroxylase (Lu et al., 2018). Tryptamine 5-hydroxylase is responsible for transforming tryptamine to serotonin and increases plant resistance against plant hoppers. A similar approach could be applied to the sugar beet plants to develop resistance against leaf chewing insects like *H. armigera* and *S. litura*. CRISPR/Cas 9-mediated mutagenesis of *GmUGT* led to the development of transgenic plant (Zhang et al., 2022). The *Arabidopsis ugt72b1* mutant exhibited aggravated cell wall lignification and an increase in flavonoid content (Lin et al., 2016). Both cell wall lignification and flavonoids contribute to resistance against leaf-chewing insects. Cell wall lignification serves as the first physical barrier against leaf-chewing insects (War et al., 2012). Hence, the enhanced resistance against leaf-chewing insects due to *GmUGT* mutations could be attributed to aggravated cell wall lignification and altered flavonoids (Zhang et al., 2022). Sugar beet plants transformed with the *cry1Ab* gene exhibited powerful defense against lepidopteron insects (Jafari et al., 2009). Sedighi et al. (2011) also reported a similar success story against an Egyptian leafworm (*Spodoptera littoralis*) infestation. Sugar beet plants have been transformed with *Cry1Ab* and *Cry1C* which exhibit resistance against the cabbage armyworm (*Spodoptera frugiperda*) (Kimoto and Shimamoto, 2001; Kimoto and Shimamoto, 2002) and on lepidopteron with *Cry1C* and *Cry2A* as well (Lytvyn et al., 2014). Regev et al. (1996) described that the *Cry1C* protein controls larvae of *Spodoptera* spp.

One of the most efficient methods for pest control is utilizing the toxic qualities of Cry protein's present in the Gram-positive pathogenic bacterium *Bacillus thuringiensis*. Different families of Cry proteins (or Bt proteins) exhibit highly selective toxicity against members of specific insect orders (Lytvyn et al., 2014). The order Lepidoptera exhibits a toxic response against proteins encoded by *Cry1* and *Cry9* genes (Table 8). Diptera and Lepidoptera show toxicity against proteins encoded by *Cry2* genes. Coleopteran insects exhibit a toxic response against proteins encoded by *Cry3* genes. Diptera also show a toxic response against the protein encoded by the *Cry2* gene (Yuan et al., 2007).

5 Conclusion and future prospects

Global laboratories are increasingly turning to CRISPR/Cas 9 editing as their instrument of choice for determining how genes work and how they might be used in other contexts. This technology is being utilized in various crop development efforts to reduce biotic and abiotic stressors. The great precision, efficacy, efficiency, cost-effectiveness, and time efficiency of editing procedures have led to their development as useful tools. The advancement in molecular tools like CRISPR/Cas

9 has opened up new approaches for genome editing in sugar beet. This technology can be used for the generation of resistant/tolerant sugar beet breeding lines/germplasm to withstand abiotic/biotic stress. In sugar beet, specific genes can be silenced or knocked out to change their functionality. The plant may benefit and adapt to the abiotic stress environment. Tolerance to such circumstances may be linked to adjustments in their physiological and biochemical mechanisms. The subsequent breeding cycles produces sugar beet cultivars that are more resilient to such challenges due to the adoption of carefully chosen tolerant breeding lines. Under drought stress conditions, this improvement becomes apparent in the plant, highlighting the plant's enhanced water usage efficiency. This can also be tested for sugar beet. Furthermore, investigations into the use of CRISPR/Cas 9 to create novel quantitative features/traits with gain-of-function mutations through replacements in sugar beet could be seen. CRISPR/Cas 9 technology promises to make a significant contribution to understanding the gene regulatory networks underlying abiotic stress response/adaptation and crop improvement initiatives to create stress-tolerant plants.

Additionally, CRISPR/Cas 9 technology applications hold great potential for addressing the challenges faced in sugar beet crops during biotic stress. Biotic stresses, resulting from pathogen and pest occurrences, significantly affect sugar beet production and yield. Traditional breeding methods have certain limitations in achieving rapid and precise genetic modification. CRISPR/Cas 9 technology has emerged as a promising solution with revolutionary approach to improve sugar beet tolerance to biotic stresses. Despite success of CRISPR/Cas 9 technology in controlling biotic stress, particularly diseases in economically significant crops, its use in insect management has not been fully utilized. Modest success has been achieved despite the intellectual exercise in creating techniques for insect pest resistance in both insects and plants. In contrast to other stresses, the main drawback has been the scarcity of target genes. Therefore, it is crucial for scientists to focus on finding sources of resistance that might serve as a foundation for insect control. To achieve this, it is necessary to evaluate the available germplasm, including wild relatives of certain crops, for pest response and to identify stress-responsive genes using multi-omics techniques. Targeted mutations turning susceptible plants into those that can control their respective pests are not far off in space or time, with such studies already in vogue. These factors, along with regulatory restrictions on gene-edited crops, may help the technique succeed in advancing not only science but also societal acceptance.

Author contributions

VM: writing, drafting, origin of the concept and compilation; AM: editing and writing, proof reading; HP: writing, compilation, concept; SS: writing and review of literature; AS: compilation of information for table. All authors contributed to the article and approved the submitted version.

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

References

- Abdelgaffar, H., Perera, O. P., and Jurat-Fuentes, J. L. (2020). ABC transporter mutations in Cry1F-resistant fall armyworm (*Spodoptera frugiperda*) do not result in altered susceptibility to selected small molecule pesticides. *Pest Manag. Sci.* 77 (2), 949–955. doi:10.1002/ps.6106
- Abo-ollo, N. A., Abdel-Rahman, M. M., Saleh, M. S., and Gohar, I. M. A. (2018). Differentiate between sugar beet (*Beta vulgaris* L) genotypes resistance to root knot nematode (*Meloidogyne incognita*) by molecular markers. *J. Agri. Chem. Biotec. Mansoura Univ.* 9 (8), 189–194. doi:10.21608/jac.2018.35234
- Abou-Elwafa, S. F., Amin, A. E. A., and Eujayl, I. (2020). Genetic diversity of sugar beet under heat stress and deficit irrigation. *Agron. J.* 112, 3579–3590. doi:10.1002/agj.2.20356
- Aglawe, S. N., Barbadikar, K. M., Mangrauthia, S. K., and Madhav, M. S. (2018). New breeding technique genome editing for crop improvement: applications, potentials and challenges. *3 Biotech* 8 (8), 336. doi:10.1007/s13205-018-1355-3
- AgResearch Magazine (2016). *New sugar beet persist curly top virus*. US department of Agriculture. <https://agresearchmag.ars.usda.gov/2016/mar/sugarbeet/>.
- Agrios, G. N. (2005). "Plant diseases caused by nematodes," in *Plant Pathology* Editor G. N. Agrios (Burlington, MA, USA: Elsevier Academic Press) 825–874. doi:10.1016/B978-0-08-047378-9.50021-X
- Ahmad, A., Khan, S. H., and Khan, Z. (2021). *CRISPR crops: the future of food security*. Springer Nature, Singapore Pte Ltd, 289. doi:10.1007/978-981-15-7142-8
- Akram, F., Sahreen, S., Aamir, F., Haq, I. U., Malik, K., Intiaz, M., et al. (2023). An insight into modern targeted genome editing technologies with a special focus on CRISPR/Cas9 and its applications. *Mole. Biotech.* 65, 227–242. doi:10.1007/s12033-022-00501-4
- Alfatih, A., Wu, J., Jan, S. U., Zhang, Z. S., Xia, J. Q., and Xiang, C. B. (2020). Loss of rice PARAQUAT TOLERANCE 3 confers enhanced resistance to abiotic stresses and increases grain yield in field. *Plant Cell Environ.* 43 (11), 2743–2754. doi:10.1111/pce.13856
- Anagholi, A., Rajabi, A., and Khayamim, S. (2018). Response of sugar beet genotypes under salinity stress in Central areas of Iran. *Int. J. Pharm. Phytopharm. Res.* 8 (6), 49–58. <https://ejippr.com/r8Be2ZB>.
- Asmaw, M., and Zawdie, B. (2021). Mechanism and applications of CRISPR/Cas9-mediated genome editing. *Biologics* 15, 353–361. doi:10.2147/BTT.S326422
- Azhar, M., Phutela, R., Kumar, M., Ansari, A. H., Rauthan, R., Gulati, S., et al. (2021). Rapid and accurate nucleobase detection using Fncas9 and its application in COVID-19 diagnosis. *Biosens. Bioelectron.* 183, 113207. doi:10.1016/j.bios.2021.113207
- Badhan, S., Ball, A. S., and Mantri, N. (2021). First report of CRISPR/Cas9 mediated DNA-free editing of 4CL and MVE7 genes in Chickpea protoplasts. *Int. J. Mol. Sci.* 22, 396. doi:10.3390/ijms22010396
- Baeg, G. J., Kim, S. H., Choi, D. M., Tripathi, S., Han, Y.-J., and Kim, J.-I. (2021). CRISPR/Cas9-mediated mutation of 5-oxoprolinase gene confers resistance to sulfonamide compounds in *Arabidopsis*. *Plant Biotechnol. Rep.* 15 (6), 753–764. doi:10.1007/s11816-021-00718-w
- Baitha, A., Srivastava, S., and Misra, V. (2022). "Insect-pests of sugar beet and their integrated management," in *Sugar Beet Cultivation, Management and Processing*. Editors V. Misra, S. Srivastava, and A. K. Mall (Singapore: Springer), 643–657. doi:10.1007/978-981-19-2730-0_31
- Bajpai, A. B., Solanki, A., Srivastava, N., Semwal, P. P., Thapliyal, P., Payal, R., et al. (2023). CRISPR-Cas System: a revolutionizing tool for genome editing. *Biochem. Cell. Arch.* 23 (1), 623–630. doi:10.51470/bca.2023.1.623
- Bakooie, M., Pourjam, E., Mahmoudi, S. B., Safaie, N., and Naderpour, M. (2015). Development of an SNP marker for sugar beet resistance/susceptible genotyping to root-knot nematode. *J. Agr. Sci. Tech.* 17, 443–454.
- Barakate, A., and Stephens, J. (2016). An overview of CRISPR-based tools and their improvements: new opportunities in understanding plant-pathogen interactions for better crop protection. *Front. Plant Sci.* 7, 765. doi:10.3389/fpls.2016.00765
- Barratt, G. E., Murchie, E. H., and Sparkes, D. L. (2023). Water use efficiency responses to fluctuating soil water availability in contrasting commercial sugar beet varieties. *Front. Plant Sci.* 14, 1119321. doi:10.3389/fpls.2023.1119321
- Barry, J. J. (2006). Root rot disease of sugar beet. *Zh. Matice Srp. za Prir. nauke*, 110. doi:10.2298/ZMSPN0610009J
- Beata, P., Roszival, M., and Kubova, V. (2022). Influence of heavy metals on growth and metabolism of sugar beet. *Lsity Cukrovanicke a Reparske* 138 (3), 112–115.
- Bhatia, S., Pooja, and Yadav, S. K. (2023). CRISPR-Cas for genome editing: classification, mechanism, designing and applications. *Int. J. Biol. Macromol.* 238, 124054. doi:10.1016/j.ijbiomac.2023.124054
- Bhattacharya, A. (2022). "Plant growth hormones in plants under low-temperature stress: a review," in *Physiological processes in plants under low temperature stress*. Editor A. Bhattacharya (Singapore: Springer), 517–627. doi:10.1007/978-981-16-9037-2_6
- Bi, H., Xu, X., Li, X., Wang, Y., Zhou, S., and Huang, Y. (2022). CRISPR/Cas9-mediated *Serine protease 2* disruption induces male sterility in *Spodoptera litura*. *Front. Physiol.* 13, 931824. doi:10.3389/fphys.2022.931824
- Bi, H. L., Xu, J., Tan, A. J., and Huang, Y. P. (2016). CRISPR/Cas9-mediated targeted gene mutagenesis in *Spodoptera litura*. *Insect Sci.* 23, 469–477. doi:10.1111/1744-7917.12341
- Bolotin, A., Quinquis, B., Sorokin, A., and Ehrlich, S. D. (2005). Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiologia* 151 (8), 2551–2561. doi:10.1099/mic.0.28048-0
- Boti, M. A., Athanasopoulou, K., Adamopoulos, P. G., Sideris, D. C., and Scorilas, A. (2023). Recent advances in genome engineering strategies. *Genes* 14 (1), 129. doi:10.3390/genes14010129
- Bozdag, G. O., Kaya, A., Koc, A., Noll, G. A., Prüfer, D., and Karakaya, H. C. (2014). Characterization of a cDNA from *Beta maritima* that confers nickel tolerance in yeast. *Gene* 538, 251–257. doi:10.1016/j.gene.2014.01.052
- Bravo, A., Gill, S. S., and Soberón, M. (2007). Mode of action of *Bacillus thuringiensis* Cry and Cyt toxins and their potential for insect control. *Toxicon* 49 (4), 423–435. doi:10.1016/j.toxicon.2006.11.022
- Briddon, R. W., Watts, J., Markham, P. G., and Stanley, J. (1989). The coat protein of beet curly top virus is essential for infectivity. *Virology* 2, 628–633. doi:10.1016/0042-6822(89)90205-5
- Buchholzer, M., and Frommer, W. B. (2023). An increasing number of countries regulate genome editing in crops. *New Phytol.* 237, 12–15. doi:10.1111/nph.18333
- Butt, H., Rao, G. S., Sedeek, K., Aman, R., Kamel, R., and Mahfouz, M. (2018). Engineering herbicide resistance via prime editing in rice. *Plant Biotechnol. J.* 18 (12), 2370–2372. doi:10.1111/pbi.13399
- Bybordi, A. (2010). Effects of salinity on yield and component characters in canola (*Brassica napus* L.) cultivars. *Not. Sci. Biol.* 2 (1), 81–83. doi:10.15835/nsb.2.1.3560
- Cai, D., Kleine, M., Kifle, S., Harloff, H. J., Sandal, N. N., Marcker, K. A., et al. (1997). Positional cloning of a gene for nematode resistance in sugar beet. *Sci* 275, 832–834. doi:10.1126/science.275.5301.832
- Carte, J., Wang, R., Li, H., Terns, R. M., and Serns, M. P. (2008). Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes Dev.* 22 (24), 3489–3496. doi:10.1101/gad.1742908
- Casarin, B. (1999). "Le annersità: loro natura, prevenzione e lotta," in *La barbabietola negli ambienti mediterranei*. Editors B. Casarin, E. Biancardi, and P. Ranalli (Bologna, Italy: Edagricole), 273–421.
- Chang, H., Liu, Y., Ai, D., Jiang, X., Dong, S., and Wang, G. (2017). A pheromone antagonist regulates optimal mating time in the moth *Helicoverpa armigera*. *Curr. Biol.* 27, 1610–1615. doi:10.1016/j.cub.2017.04.035
- Chang, J. D., Huang, S., Yamaji, N., Zhang, W., Ma, J. F., and Zhao, F. J. (2020). OsNRAMP1 transporter contributes to cadmium and manganese uptake in rice. *Plant Cell Environ.* 43, 2476–2491. doi:10.1111/pce.13843
- Chaturvedi, S. (2004). Biosafety regulation: need for fine balancing. *Econ. Polit. Wkly.* 39, 3693–3697.
- Chaudhuri, A., Halder, K., and Datta, A. (2022). Classification of CRISPR/Cas system and its application in tomato breeding. *Theor. Appl. Genet.* 135, 367–387. doi:10.1007/s00122-021-03984-y
- Chen, J. S., Ma, E., Harrington, L. B., Costa, M. D., Tian, X., Palefsky, M., et al. (2018a). CRISPR-Cas12a target binding unleashes indiscriminate single stranded DNase activity. *Sci* 360 (6387), 436–439. doi:10.1126/science.aar6245
- Chen, K., and Gao, C. (2020). Genome edited crops: how to move them from laboratory to market. *Front. Agr. Sci. Eng.* 7 (2), 181–187. doi:10.15302/J-FASE-2020332
- Chen, L., Wei, J., Liu, C., Zhang, W., Wang, B., Niu, L., et al. (2018b). Specific binding protein ABC1 is associated with Cry2Ab toxicity in *Helicoverpa armigera*. *Front. Physiol.* 19 (9), 745. doi:10.3389/fphys.2018.00745

- Chen, X., Cao, Y., Zhan, S., Tan, A., Palli, S. R., and Huang, Y. (2019). Disruption of sex-specific double sex exons results in male- and female-specific defects in the black cutworm, *Agrotis ipsilon*. *Pest Manag. Sci.* 75 (6), 1697–1706. doi:10.1002/ps.5290
- Chen, X., Cao, Y., Zhan, S., Zhang, Y., Tan, A., and Huang, Y. (2018c). Identification of yellow gene family in *Agrotis ipsilon* and functional analysis of Aiyellow-y by CRISPR/Cas 9. *Insect biochem. Mol. Biol.* 94, 1–9. doi:10.1016/j.ibmb.2018.01.002
- Chinnusamy, V., Zhu, J., and Zhu, J. K. (2006). Gene regulation during cold acclimation in plants. *Physiol. Plant.* 126 (1), 52–61. doi:10.1111/j.1399-3054.2006.00596.x
- Chu, C., Huang, R., Liu, L., Tang, G., Xiao, J., Yoo, H., et al. (2022). The rice heavy-metal transporter OsNRAMP1 regulates disease resistance by modulating ROS homeostasis. *Plant Cell Environ.* 45, 1109–1126. doi:10.1111/pce.14263
- Collinge, D. B., Kragh, K. M., Mikkelsen, J. D., Nielsen, K. K., Rasmussen, U., and Vad, K. (1993). Plant chitinases. *Plant J.* 3 (1), 31–40. doi:10.1046/j.1365-313x.1993.t01-1-00999.x
- Cooke, D. A. (1987). Beet cyst nematode (*Heterodera schachtii* Schmidt) and its control on sugar beet. *Agricul. Zool. Rev.* 2, 135–183.
- Cui, Y., Sun, J. L., and Yu, L. (2017). Application of the CRISPR gene-editing technique in insect functional genome studies - a review. *Entomol. Exp. Appl.* 162 (2), 124–132. doi:10.1111/eea.12530
- Davis, C. A., Hitz, B. C., Sloan, C. A., Chan, E. T., Davidson, J. M., Gabdank, I., et al. (2018). The encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res.* 46 (1), D794–D801. doi:10.1093/nar/gkx1081
- Deveau, H., Barrangou, R., Garneau, J. E., Labonté, J., Fremaux, C., Boyaval, P., et al. (2008). Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J. Bacteriol.* 190 (4), 1390–1400. doi:10.1128/JB.01412-07
- DiFonzo, C., Jewett, M., Warner, F., Brown-Rytlewski, D., and Kirk, W. (2006). Insect, nematode, and disease control in Michigan field crops. MSU Bulletin E-1582 2006. Michigan State University East Lansing, MI 48824 https://www.canr.msu.edu/field_crops/uploads/archive/Part18E1582DryBeanDiseases.pdf.
- Dobrovidova, O. (2019). Russia joins in global gene-editing bonanza. *Nature* 569, 319–320. doi:10.1038/d41586-019-01519-6
- Du, Y. T., Zhao, M. J., Wang, C. T., Gao, Y., Wang, Y. X., Liu, Y. W., et al. (2018). Identification and characterization of GmMYB118 responses to drought and salt stress. *BMC Plant Biol.* 18, 320. doi:10.1186/s12870-018-1551-7
- Duan, Y. B., Li, J., Qin, R. Y., Xu, R. F., Li, H., Yang, Y. C., et al. (2016). Identification of a regulatory element responsible for salt induction of rice OsRAV2 through ex-situ and in situ promoter analysis. *Plant Mol. Biol.* 90, 49–62. doi:10.1007/s11103-015-0393-z
- Dutta, T. K., Papolu, P. K., Banakar, P., Chiudhary, D., Sirohi, A., and Rao, U. (2015). Tomato transgenic plants expressing hairpin construct of a nematode protease gene conferred enhanced resistance to root knot nematodes. *Front. Microbiol.* 6, 260. doi:10.3389/fmicb.2015.00260
- Ebrahimi, S., Alexandra Babler, A., Eini, O., Yildirim, Z., Wassenegger, W., Krczal, G., et al. (2022). Beet curly top Iran virus Rep and V2 gene work as silencing suppressors through separate mechanisms. *bioRxiv*. doi:10.1101/2022.08.25.505242
- Eini, O., Schumann, N., Niessen, M., and Varrelmann, M. (2022). Targeted mutagenesis in plants using Beet curly top virus for efficient delivery of CRISPR/Cas12a components. *New Biotech.* 67, 1–11. doi:10.1016/j.nbt.2021.12.002
- Erbasol, I., Bozdogan, G. O., Koc, A., Pedas, P., and Karakaya, H. C. (2013). Characterization of two genes encoding metal tolerance proteins from *Beta vulgaris* subspecies maritima that confers manganese tolerance in yeast. *BioMetals* 26, 795–804. doi:10.1007/s10534-013-9658-7
- Esh, A., and Taghian, S. (2022). “Etiology, epidemiology, and management of sugar beet diseases,” in *Sugar beet cultivation, management and processing*. Editors V. Misra, S. Srivastava, and A. K. Mall (Singapore: Springer), 505–540. doi:10.1007/978-981-19-2730-0_25
- Fan, K., Mao, Z., Ye, F., Pan, X. F., Li, Z., Lin, W., et al. (2021). Genome-wide identification and molecular evolution analysis of the heat shock transcription factor (HSF) gene family in four diploid and two allopolyploid *Gossypium* species. *Genomics* 113 (5), 3112–3127. doi:10.1016/j.ygeno.2021.07.008
- Fang, S., Hou, X., and Liang, X. (2021). Response mechanisms of plants under saline-alkali stress. *Front. Plant Sci.* 12, 667458. doi:10.3389/fpls.2021.667458
- Fang, Y., and Xiong, L. (2015). General mechanisms of drought response and their application in drought resistance improvement in plants. *Cell Mol. Life Sci.* 72 (4), 673–689. doi:10.1007/s00018-014-1767-0
- FAO (2022). *Agriculture statistics*. Rome, Italy: Food and Agriculture Organization of the United Nations.
- Fineran, P. C., Gerritzen, M. J., Suárez-Díez, M., Künne, T., Boekhorst, J., van Hijum, S. A., et al. (2014). Degenerate target sites mediate rapid primed CRISPR adaptation. *PNAS* 111 (16), E1629–E1638. doi:10.1073/pnas.1400071111
- Flagel, L., Lee, Y. W., Wanjugi, H., Swarup, S., Brown, A., Wang, J., et al. (2018). Mutational disruption of the ABCC2 gene in fall armyworm *Spodoptera frugiperda* confers resistance to the Cry1Fa and Cry1Aa105 insecticidal proteins. *Sci. Rep.* 8, 7255. doi:10.1038/s41598-018-25491-9
- Francis, S. (2002). Sugar beet powdery mildew (*Erysiphe betae*). *Mol. Plant Pathol.* 3 (3), 119–124. doi:10.1046/j.1364-3703.2002.00103.x
- Francis, S. A., and Luterbacher, M. C. (2003). Identification and exploitation of novel disease resistance genes in sugar beet. *Pest Manag. Sci.* 59, 225–230. doi:10.1002/ps.569
- Garg, R., Verma, M., Agrawal, S., Shankar, R., Majee, M., and Jain, M. (2014). Deep transcriptome sequencing of wild halophyte rice, *Porteresia coarctata*, provides novel insights into the salinity and submergence tolerance factors. *DNA Res.* 21, 69–84. doi:10.1093/dnares/dst042
- Ghaemi, R., Pourjam, E., Safaie, N., Verstraeten, B., Mahmoudi, S. B., Mehrabi, R., et al. (2020). Molecular insights into the compatible and incompatible interactions between sugar beet and the beet cyst nematode. *BMC Plant Biol.* 20, 483. doi:10.1186/s12870-020-02706-8
- Ghaffari, H., Tadayon, M. R., Bahador, M., and Razmjoo, J. (2022). Biochemical and yield response of sugar beet to drought stress and foliar application of vermicompost tea. *Plant Stress* 5, 100087. doi:10.1016/j.stress.2022.100087
- Ghouri, M. Z., Munawar, N., Aftab, S. O., and Ahmad, A. (2023). “Regulation of CRISPR edited food and feed: legislation and future,” in *GMOs and political stance*. Editors M. A. Nawaz, K. S. Golokhvast, G. Chung, and A. M. Tsatsakis (Apple Academic Press), 261–287.
- Gidner, S., Lennfors, B. L., Nilsson, N. O., Bensefelt, J., Johansson, E., Gyllenspetz, U., et al. (2005). QTL mapping of BNYVV resistance from the WB41 source in sugar beet. *Genome* 48, 279–285. doi:10.1139/g04-108
- Gleditsch, D., Pausch, P., Muller-Esparza, H., Ozcan, A., Guo, X., Bange, G., et al. (2019). PAM identification by CRISPR-Cas effector complexes: diversified mechanisms and structures. *RNA Biol.* 16 (4), 504–517. doi:10.1080/15476286.2018.1504546
- Gootenberg, J. S., Abudayyeh, O. O., Kellner, M. J., Joung, J., Collins, J. J., and Zhang, F. (2018). Multiplexed and portable nucleic acid detection platform with Cas13, Cas12a and Csm6. *Sci.* 80 (360), 439–444. doi:10.1126/science.aag0179
- Gootenberg, J. S., Abudayyeh, O. O., Lee, J. W., Essletzbichler, P., Dy, A. J., Joung, J., et al. (2017). Nucleic acid detection with CRISPR-Cas13a/C2c2. *Sci.* 80 (356), 438–442. doi:10.1126/science.aam9321
- Greco, N., D’Addabbo, T., Brandonisio, A., and Elia, F. (1993). Damage to Italian crops caused by cyst forming nematodes. *J. Nematol.* 25, 836–842. PMID: PMC2619465.
- Greger, M., and Ögren, E. (1991). Direct and indirect effects of Cd²⁺ on photosynthesis in sugar beet (*Beta vulgaris*). *Physiol. Plant* 83, 129–135. doi:10.1111/j.1399-3054.1991.tb01291.x
- Grimmer, M. K., Bean, K. M. R., and Asher, M. J. C. (2007). Mapping of five resistance genes to sugar-beet powdery mildew using AFLP and anchored SNP markers. *Theor. Appl. Genet.* 115 (1), 67–75. doi:10.1007/s00122-007-0541-1
- Grujicic, G. (1958). *Heterodera schachtii* Schmidt a beet nematode in our country. *Plant Prot.* 49 (50), 167–174.
- Guo, M., Liu, J. H., Ma, X., Luo, D. X., Gong, Z. H., and Lu, M. H. (2016). The plant heat stress transcription factors (HSFs): structure, regulation, and function in response to abiotic stresses. *Front. Plant Sci.* 7, 114. doi:10.3389/fpls.2016.00114
- Gupta, N., Reddy, K. K., Bhattacharya, D., and Chakraborty, S. (2021). Plant responses to geminivirus infection: guardians of the plant immunity. *Virology* 18, 143. doi:10.1186/s12985-021-01612-1
- Haq, A. F. M., Tasnim, J., El-Shehawi, A. M., Rahman, M. A., Parvez, M. S., Ahmed, M. B., et al. (2021). The Cd-induced morphological and photosynthetic disruption is related to the reduced Fe status and increased oxidative injuries in sugar beet. *Plant Physiol. Biochem.* 166, 448–458. doi:10.1016/j.plaphy.2021.06.020
- Haq, E., Taniguchi, H., Hassan, M., Bhowmik, P., Karim, M. R., Smiech, M., et al. (2018). Application of CRISPR/Cas9 genome editing technology for the improvement of crops cultivated in tropical climates: recent progress, prospects, and challenges. *Front. Plant Sci.* 9, 617. doi:10.3389/fpls.2018.00617
- Harveson, R. M. (2013). Cercospora leaf spot of sugar beet. Extension bulletin Available at: <https://extensionpublications.unl.edu/assets/pdf/g1753.pdf>.
- Haurwitz, R. E., Jinek, M., Wiedenheft, B., Zhou, K., and Doudna, J. A. (2010). Sequence and structure-specific RNA processing by a CRISPR endonuclease. *Sci* 329 (5997), 1355–1358. doi:10.1126/science.1192272
- Hennig, L. (2012). Plant gene regulation in response to abiotic stress. *Biochim. Biophys. Acta. Gen. Subj.* 1819 (2), 85. doi:10.1016/j.bbagr.2012.01.005
- Hernández-Rodríguez, C. S., Hernández-Martínez, P., Rie, J. V., Escribano, B., and Ferré, J. (2013). Shared midgut binding sites for Cry1A.105, Cry1Aa, Cry1Ab, Cry1Ac and Cry1Fa proteins from *Bacillus thuringiensis* in two important corn pests, *Ostrinia nubilalis* and *Spodoptera frugiperda*. *PloS One* 8 (7), e68164. doi:10.1371/journal.pone.0068164
- Herrera, F. F., Domené-Painena, O., and Cruces, J. M. (2017). The history of agroecology in Venezuela: a complex and multifaceted process. *Agroecol. Sustain. Food Syst.* 41, 401–415. doi:10.1080/21683565.2017.1285842
- Hoffmann, C., and Kluge-Severin, S. (2011). Growth analysis of autumn and spring sown sugar beet. *Eur. J. Agron.* 34, 1–9. doi:10.1016/j.eja.2010.09.001

- Horlbeck, M. A., Gilbert, L. A., Villalta, J. E., Adamson, B., Pak, R. A., Chen, Y., et al. (2016). Compact and highly active next generation libraries for CRISPR mediated gene repression and activation. *Elife* 5, e19760. doi:10.7554/eLife.19760
- Hsu, P. D., Lander, E. S., and Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. *Cell* 157, 1262–1278. doi:10.1016/j.cell.2014.05.010
- Huanga, J., Xua, Y., Zuoa, Y., Yanga, Y., Tabashnikb, B. E., and Wua, Y. (2020). Evaluation of five candidate receptors for three Bt toxins in the beet armyworm using CRISPR-mediated gene knockouts. *Insect biochem. Mol. Biol.* 121, 103361. doi:10.1016/j.ibmb.2020.103361
- Iqbal, Z., Memom, A. G., Ahmad, A., and Iqbal, M. S. (2022). Calcium mediated cold acclimation in plants: underlying signaling and molecular mechanisms. *Front. Plant Sci.* 13, 855559. doi:10.3389/fpls.2022.855559
- Ismail, R. M., Youssef, A. B., El-Assal, S. E. D., Tawfik, M. S., and Abdallah, N. A. (2020). Cloning and characterization of heat shock factor (BVHSF) from sugar beet (*Beta vulgaris* L.). *Plant Arch.* 20 (2), 3725–3733.
- Jacob, P., Hirt, H., and Bendahmane, A. (2017). The heat-shock protein/chaperone network and multiple stress resistance. *Plant Biotechnol. J.* 15, 405–414. doi:10.1111/pbi.12659
- Jafari, M., Norouzi, P., Malboobi, M. A., Ghareyazie, B., Valizadeh, M., Mohammadi, S. A., et al. (2009). Enhanced resistance to a lepidopteran pest in transgenic sugar beet plants expressing synthetic cry1Ab gene. *Euphytica* 165, 333–344. doi:10.1007/s10681-008-9792-4
- Jalilian, M., Dehdari, M., Fahlani, R. R., and Dehnovi, M. M. (2017). Study of cold tolerance of different sugar beet (*Beta vulgaris* L.) cultivars at seedling growth stage. *Environ. Stresses Crop Sci.* 10, Pe475–Pe490. doi:10.22077/ESCS.2017.616,
- Jamla, M., Khare, T., Joshi, S., Patil, S., Penna, S., and Kumar, V. (2021). Omics approaches for understanding heavy metal responses and tolerance in plants. *Curr. Plant Biol.* 27, 100213. doi:10.1016/j.cpb.2021.100213
- Ji, X., Zhang, H., Zhang, Y., Wang, Y., and Gao, C. (2015). Establishing a CRISPR-Cas-like immune system conferring DNA virus resistance in plants. *Nat. Plants* 1, 15144. doi:10.1038/nplants.2015.144
- Jiang, J., Ma, S., Ye, N., Jiang, M., Cao, J., and Zhang, J. (2017). WRKY transcription factors in plant responses to stresses. *J. Integr. Plant Biol.* 59 (2), 86–101. doi:10.1111/jipb.12513
- Jiang, N. M., Zhang, C., Liu, J. Y., Guo, Z. H., Zhang, Z. Y., Han, C. G., et al. (2019). Development of Beet necrotic yellow vein virus-based vectors for multiple-gene expression and guide RNA delivery in plant genome editing. *Plant Biotech. J.* 17, 1302–1315. doi:10.1111/pbi.13055
- Jiang, S. Y., Ramamoorthy, R., Bhalla, R., Luan, H. F., Venkatesh, P. N., Cai, M., et al. (2008). Genome-wide survey of the RIP domain family in *Oryza sativa* and their expression profiles under various abiotic and biotic stresses. *Plant Mol. Biol.* 67, 603–614. doi:10.1007/s11103-008-9342-4
- Jiang, W., Bikard, D., Cox, D., Zhang, F., and Marraffini, L. A. (2013). RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat. Biotechnol.* 31 (3), 233–239. doi:10.1038/nbt.2508
- Jiang, Y., and Deyholos, M. K. (2008). Functional characterization of *Arabidopsis* NaCl-inducible WRKY25 and WRKY33 transcription factors in abiotic stresses. *Plant Mol. Biol.* 69, 91–105. doi:10.1007/s11103-008-9408-3
- Jiao, C., Sharma, S., Dugar, G., Peeck, N. L., Bischler, T., Wimmer, F., et al. (2021). Noncanonical crRNAs derived from host transcripts enable multiplexable RNA detection by Cas9. *Sci.* 80 (372), 941–948. doi:10.1126/science.abe7106
- Jin, L., Wang, J., Guan, F., Zhang, J., Yu, S., Liu, S., et al. (2018). Dominant point mutation in a tetraspanin gene associated with field-evolved resistance of cotton bollworm to transgenic Bt cotton. *Proc. Natl. Acad. Sci. U. S. A.* 115, 11760–11765. doi:10.1073/pnas.1812138115
- Jin, M., Tao, J., Li, Q., Cheng, Y., Sun, X., Wu, K., et al. (2021). Genome editing of the SfABCC2 gene confers resistance to Cry1F toxin from *Bacillus thuringiensis* in *Spodoptera frugiperda*. *J. Integr. Agric.* 20, 815–820. doi:10.1016/S2095-3119(19)62772-3
- Jin, M., Yang, Y., Shan, Y., Chakrabarty, S., Cheng, Y., Soberon, M., et al. (2020). Two ABC transporters are differentially involved in the toxicity of two *Bacillus thuringiensis* Cry1 toxins to the invasive crop-pest *Spodoptera frugiperda* (J. E. Smith). *Pest Manag. Sci.* 77, 1492–1501. doi:10.1002/ps.6170
- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Sci.* 337 (6096), 816–821. doi:10.1126/science.1225829
- Johansson, E. (1985). Rhizomania in sugar beet—a threat to beet growing that can be overcome by plant breeding. *Sveriges Utsädesförenings Tidskr.* 95, 115–121.
- Joshi, I., Kumar, A., Kohli, D., Singh, A. K., Sirohi, A., Subramaniam, K., et al. (2020). Conferring root-knot nematode resistance via host-delivered RNAi-mediated silencing of four Mi-msp genes in *Arabidopsis*. *Plant Sci.* 298, 110592. doi:10.1016/j.plantsci.2020.110592
- Jyoti, K., Meenakshi, F., and Neetika, M. (2023). CRISPR/Cas9: an evolutionary approach towards crops amelioration. *Int. J. Life Sci.* 12 (1), 44–61. doi:10.5958/2319-1198.2023.00004.0
- Kalaitzandonakes, N., Willig, C., and Zahringer, K. (2022). The economics and policy of genome editing in crop improvement. *Plant Genome* 16 (2), e20248. doi:10.1002/tpg2.20248
- Karim, S., Riazuddin, S., Gould, F., and Dean, D. H. (2000). Determination of receptor binding properties of *Bacillus thuringiensis* delta-endotoxins to cotton bollworm (*Helicoverpa zea*) and pink bollworm (*Pectinophora gossypiella*) midgut brush border membrane vesicles. *Pestic. Biochem. Physiol.* 67, 198–216. doi:10.1006/pest.2000.2491
- Karmakar, S., Das, P., Panda, D., Xie, K., Baig, M. J., and Molla, K. A. (2022). A detailed landscape of CRISPR Cas mediated plant disease and pest management. *Plant Sci.* 323, 111376. doi:10.1016/j.plantsci.2022.111376
- Keller, I., Müdsam, C., Martins Rodrigues, C., Kischka, D., Zierer, W., Sonnewald, U., et al. (2021). Cold-triggered induction of ROS- and raffinose metabolism in freezing-sensitive taproot tissue of sugar beet. *Front. Plant Sci.* 12, 715767. doi:10.3389/fpls.2021.715767
- Khadiza, K., Khan, R. A. H., Park, J. I., Nath, U. K., Kimm, K. B., Nou, I. S., et al. (2017). Molecular characterization and expression profiling of tomato GRF transcription factor family genes in response to abiotic stresses and phytohormones. *Int. J. Mole. Sci.* 18 (5), 1056. doi:10.3390/ijms18051056
- Khan, S. A., Reichelt, M., and Heckel, D. G. (2017). Functional analysis of the ABCs of eye color in *Helicoverpa armigera* with CRISPR/Cas9-induced mutations. *Sci. Rep.* 7, 40025. doi:10.1038/srep40025
- Khatodia, S., Bhatotia, K., and Tuteja, N. (2017). Development of CRISPR/Cas9 mediated virus resistance in agriculturally important crops. *Bioengineered* 8 (3), 274–279. doi:10.1080/21655979.2017.1297347
- Kim, S. T., Choi, M., Bae, S. J., and Kim, J. S. (2021). The functional association of acqos/victr with salt stress resistance in *Arabidopsis thaliana* was confirmed by CRISPR-mediated mutagenesis. *Int. J. Mol. Sci.* 22, 11389. doi:10.3390/ijms222111389
- Kimoto, Y., and Shimamoto, Y. (2001). Difference in toxicity to larvae of cabbage armyworm between transgenic sugar beet lines with Cry1Ab and Cry1AC. *Proc. J. Soc. Sugar Beet Technol.* 43, 20–23.
- Kimoto, Y., and Shimamoto, Y. (2002). Differences in toxicity to larvae of cabbage armyworm between transgenic sugar beet lines with cry I A(b) and cry I C. *Proc. J. Soc. Sugar Beet Technol.* 43, 20–23.
- Kito, K., Yamane, K., Yamamori, T., Matsuhira, H., Tanaka, Y., and Takabe, T. (2018). Isolation, functional characterization and stress responses of raffinose synthase genes in sugar beet. *J. Plant Biochem. Biotechnol.* 27, 36–45. doi:10.1007/s13562-017-0413-y
- Kocak, D., and Gersbach, C. (2018). From CRISPR scissors to virus sensors. *Nature* 557 (7704), 168–169. doi:10.1038/d41586-018-04975-8
- Koutroumpa, F. A., Monsempe, C., Francois, M. C., Cian, A. D., Royer, C., Concordet, J. P., et al. (2016). Heritable genome editing with CRISPR/Cas9 induces anosmia in a crop pest moth. *Sci. Rep.* 6, 29620. doi:10.1038/srep29620
- Kumar, A., Harloff, H. J., Melzer, S., Leineweber, J., Defant, B., and Jung, C. (2021). A rhomboid-like protease gene from an interspecies translocation confers resistance to cyst nematodes. *New Phytol.* 231, 801–813. doi:10.1111/nph.17394
- Kumar, A., Yang, F., Goddard, L., and Schubert, S. (2004). Differing trends in the tropical surface temperatures and precipitation over land and oceans. *J. Clim.* 17, 653–664. doi:10.1175/1520-0442(2004)017<0653:DTTTS>2.0.CO;2
- Kumar, K., Gambhir, G., Dass, A., Tripathi, A. K., Singh, A., Jha, A. K., et al. (2020). Genetically modified crops: current status and future prospects. *Planta* 251, 91. doi:10.1007/s00425-020-03372-8
- Lan, T., Zheng, Y., Su, Z., Yu, S., Song, H., Zheng, X., et al. (2019). OsSPL10, a SBP-Box gene, plays a dual role in salt tolerance and trichome formation in rice (*Oryza sativa* L.). *G3 Genes, Genomes, Genet.* 9 (12), 4107–4114. doi:10.1534/g3.119.400700
- Larbi, A., Morales, F., Abadía, A., Gogorcena, Y., Lucena, J., and Abadía, J. (2002). Effects of Cd and Pb in sugar beet plants grown in nutrient solution: induced Fe deficiency and growth inhibition. *Funct. Plant Biol.* 29, 1453–1464. doi:10.1071/FP02090
- La Russa, M. F., and Qi, L. S. (2015). The new state of the art: cas9 for gene activation and repression. *Mol. Cell Biol.* 35 (22), 3800–3809. doi:10.1128/MCB.00512-15
- Lei, G., Shen, M., Li, Z. G., Zhang, B., Duan, K. X., Wang, N., et al. (2020). EIN2 regulates salt stress response and interacts with a MA3 domain-containing protein ECIP1 in *Arabidopsis*. *Plant Cell Environ.* 34 (10), 1678–1692. doi:10.1111/j.1365-3040.2011.02363.x
- Lein, J. C., Asbach, K., Tian, Y., Schulte, D., Li, C., Koch, G., et al. (2007). Resistance gene analogues are clustered on chromosome 3 of sugar beet and cosegregate with QTL for rhizomania resistance. *Genome* 50, 61–71. doi:10.1139/g06-131
- Lewellen, R. T., and Schrandt, J. K. (2001). Inheritance of powdery mildew resistance in sugar beet derived from *Beta vulgaris* subsp. *maritima*. *Plant Dis.* 85, 627–631. doi:10.1094/PDIS.2001.85.6.627
- Lewellen, R. T., Skoyen, I. O., and Erichsen, A. W. (1987). “Breeding sugar beet for resistance to rhizomania: evaluation of host-plant reactions and selection for and inheritance of resistance,” in Proceedings of the IIRB 50th Congress, 139–156.
- Li, B., Liang, S., Alariqi, M., Wang, F., Wang, G., Wang, Q., et al. (2021a). The application of temperature sensitivity CRISPR/LbCpf1 (LbCas12a) mediated genome

editing in allotetraploid cotton (*G. Hirsutum*) and creation of non-transgenic, gossypol-free cotton. *Plant Biotechnol. J.* 19 (2), 221–223. doi:10.1111/pbi.13470

Li, H., Xu, Y., Xiao, Y., Zhu, Z., Xie, X., Zhao, H., et al. (2010). Expression and functional analysis of two genes encoding transcription factors, VpWRKY1 and VpWRKY2, isolated from Chinese wild *Vitis pseudoreticulata*. *Planta* 232, 1325–1337. doi:10.1007/s00425-010-1258-y

Li, J., Cui, J., Dai, C., Liu, T., Cheng, D., and Luo, C. (2021b). Whole-transcriptome RNA sequencing reveals the global molecular responses and cerna regulatory network of mRNAs, lncRNAs, miRNAs and circRNAs in response to salt stress in sugar beet (*Beta vulgaris*). *Int. J. Mol. Sci.* 22, 289. doi:10.3390/ijms22010289

Li, J., Wang, Y., Wang, B., Lou, J., Ni, P., Jin, Y., et al. (2022a). Application of CRISPR/Cas systems in the nucleic acid detection of infectious diseases. *Diagnostics* 12, 2455. doi:10.3390/diagnostics12102455

Li, J., Zhang, M., Sun, J., Mao, X., Wang, J., Liu, H., et al. (2020a). Heavy metal stress associated proteins in rice and *Arabidopsis*: genome-wide identification phylogenetics, duplication and expression profile analysis. *Front. Gene.* 11, 477. doi:10.3389/fgene.2020.00477

Li, Q., Jin, M., Yu, S., Cheng, Y., Shan, Y., Wang, P., et al. (2022b). Knockout of the ABCB1 gene increases susceptibility to emamectin benzoate, beta-cypermethrin and chlorantraniliprole in *Spodoptera frugiperda*. *Insects* 13, 137. doi:10.3390/insects13020137

Li, R., Zhang, L., Wang, L., Chen, L., Zhao, R., Sheng, J., et al. (2018). Reduction of tomato-plant chilling tolerance by CRISPR–Cas9 mediated SlCBF1 mutagenesis. *J. Agric. Food Chem.* 66, 9042–9051. doi:10.1021/acs.jafc.8b02177

Li, S., Fu, Q., Chen, L., Huang, W.-D., and Yu, D. (2011). *Arabidopsis thaliana* WRKY25, WRKY26, and WRKY33 coordinate induction of plant thermotolerance. *Planta* 233, 1237–1252. doi:10.1007/s00425-011-1375-2

Li, W., Pang, S., Lu, Z., and Jin, B. (2020b). Function and mechanism of WRKY transcription factors in abiotic stress responses of plants. *Plants* 9, 1515. doi:10.3390/plants9111515

Liao, S., Qin, X., Luo, L., Han, Y., Wang, X., Usman, B., et al. (2019). CRISPR/Cas9-induced mutagenesis of *semi-rolled Leaf1*, 2 confers curled leaf phenotype and drought tolerance by influencing protein expression patterns and ROS scavenging in rice (*Oryza sativa* L.). *Agron.* 9, 728. doi:10.3390/agronomy9110728

Liebe, S., Wibberg, D., Maiss, E., and Varrelmann, M. (2020). Application of a reverse genetic system for beet necrotic yellow vein virus to study Rz1 resistance response in sugar beet. *Front. Plant Sci.* 10, 1703. doi:10.3389/fpls.2019.01703

Lim, C., Kang, K., Shim, Y., Yoo, S. C., and Paek, N. C. (2022). Inactivating transcription factor OsWRKY5 enhances drought tolerance through abscisic acid signaling pathways. *Plant Physiol.* 188 (4), 1900–1916. doi:10.1093/plphys/kiab492

Lin, J. S., Huang, X. X., Li, Q., Cao, Y., Bao, Y., Meng, X. F., et al. (2016). UDP glycosyl transferase 72B1 catalyzes the glucose conjugation of monolignols and is essential for the normal cell wall lignification in *Arabidopsis thaliana*. *Plant J.* 88, 26–42. doi:10.1111/tj.13229

Liu, D., Gao, Z., Li, J., Yao, Q., Tan, W., Xing, W., et al. (2022). Effects of cadmium stress on the morphology, physiology, cellular ultrastructure, and BvHIPP24 gene expression of sugar beet (*Beta vulgaris* L.). *Int. J. Phytoremediation* 25 (4), 455–465. doi:10.1080/15226514.2022.2090496

Lombardo, L., and Grando, M. S. (2020). Genetically modified plants for nutritionally improved food: a promise kept? *Food Rev. Int.* 36, 58–76. doi:10.1080/87559129.2019.1613664

Lou, D., Wang, H., Liang, G., and Yu, D. (2017). OsSAPK2 confers abscisic acid sensitivity and tolerance to drought stress in rice. *Front. Plant Sci.* 8, 993. doi:10.3389/fpls.2017.00993

Lu, H. P., Liu, S. M., Xu, S. L., Chen, W.-Y., Zhou, X., Tan, Y.-Y., et al. (2017). CRISPR-S: an active interference element for a rapid and inexpensive selection of genome-edited, transgene-free rice plants. *Plant Biotechnol. J.* 15 (11), 1371–1373. doi:10.1111/pbi.12788

Lu, H. P., Luo, T., Fu, H. W., Wang, L., Tan, Y. Y., Huang, J. Z., et al. (2018). Resistance of rice to insect pests mediated by suppression of serotonin biosynthesis. *Nat. Plants* 4, 338–344. doi:10.1038/s41477-018-0152-7

Luterbacher, M. C., Smith, J. M., and Asher, M. J. C. (1998). Sources of disease resistance in wild Beta germplasm. *Asp. Appl. Biol.* 52, 423–430.

Luterbacher, M. C., Smith, J. M., Asher, M. J. C., and Frese, L. (2000). Disease resistance in collections of Beta species. *J. Sugar Beet Res.* 37, 39–47. doi:10.5274/jsbr.37.3.39

Lytvyn, D. I., Syvura, V. V., Kurylo, V. V., Olenieva, V. D., Yemets, A. I., and Blume, Y. B. (2014). Creation of transgenic sugar beet lines expressing insect pest resistance genes cry1C and cry2A. *Cytol. Genet.* 48 (2), 69–75. doi:10.3103/S0095452714020078

Ly, Y.-S., Cao, L.-M., Huang, W.-Q., Liu, J.-X., and Lu, H.-P. (2022). Disruption of three polyamine uptake transporter genes in rice by CRISPR/Cas9 gene editing confers tolerance to herbicide paraquat. *aBIOTECH* 3 (2), 140–145. doi:10.1007/s42994-022-00075-4

Majumdar, R., Galewski, P. J., Eujayl, I., Minocha, R., Vincill, E., and Strausbaugh, C. A. (2022). Regulatory roles of small non-coding RNAs in sugar beet resistance against beet curly top virus. *Front. Plant Sci.* 12, 780877. doi:10.3389/fpls.2021.780877

Mall, A. K., Misra, V., SanteshwariPathak, A. D., and Srivastava, S. (2022b). Sugar beet cultivation in India: prospects for bioethanol production and value-added co-products. *Sugar Tech.* 23, 1218–1234. doi:10.1007/s12355-021-01007-0

Mall, A. K., Misra, V., Srivastava, S., and Pathak, A. D. (2022a). “India’s sugar beet seed technology and production,” in *Sugar beet cultivation, management and processing*. Editors V. Misra, S. Srivastava, and A. K. Mall (Springer Nature Singapore Pte Ltd), 121–129. doi:10.1007/978-981-19-2730-0_7

Mallapaty, S. (2022). China’s approval of gene-edited crops energizes researchers. *Nature* 602, 559–560. doi:10.1038/d41586-022-00395-x

Mao, Y., Botella, J. R., Liu, Y., and Zhu, J.-K. (2019). Gene editing in plants: progress and challenges. *Natl. Sci. Rev.* 6 (3), 421–437. doi:10.1093/nsr/nwz005

McCarty, N. S., Graham, A. E., Studena, L., and Ledesma-Amaro, R. (2020). Multiplexed CRISPR technologies for gene editing and transcriptional regulation. *Nat. Commun.* 11, 1281. doi:10.1038/s41467-020-15053-x

Miao, C., Xiao, L., Hua, K., Zou, C., Zhao, Y., Bressan, R. A., et al. (2018). Mutations in a subfamily of abscisic acid receptor genes promote rice growth and productivity. *Proc. Natl. Acad. Sci. U.S.A.* 115, 6058–6063. doi:10.1073/pnas.1804774115

Mickelbart, M. V., Hasegawa, P. M., and Bailey-Serres, J. (2015). Genetic mechanisms of abiotic stress tolerance that translate to crop yield stability. *Nat. Rev. Genet.* 16, 237–251. doi:10.1038/nrg3901

Misra, V., Mall, A. K., Kumar, A., Srivastava, S., and Pathak, A. D. (2020a). Identification of two new *Alternaria* isolates on sugar beet (*Beta vulgaris* L.) plants in Lucknow, India. *Arch. Phytopathol. Plant Prot.* 54, 164–176. doi:10.1080/03235408.2020.1824378

Misra, V., Mall, A. K., Kumar, M., Singh, D., and Pathak, A. D. (2018). “Sugar beet crop: asset for farmers in enhancing income India international science festival,” in *Theme frontier areas in science (book 3) held at indra gandhi pratishtan* (Lucknow), 43. Abstract No. 43.

Misra, V., Mall, A. K., and Pathak, A. D. (2020b). “Sugar beet: a sustainable crop for saline environment,” in *Agronomic crops*. Editor M. Hassanzuman (Springer Nature Singapore Pvt. Ltd), 49–61. doi:10.1007/978-981-15-0025-1_4

Misra, V., Srivastava, S., and Mall, A. K. (2022a). *Sugar beet cultivation, management and processing*. Springer Nature Singapore Pte Ltd, 1–1005. doi:10.1007/978-981-19-2730-0

Misra, V., Srivastava, S., Mall, A. K., and Srivastava, S. (2022c). “Foliar sugar beet diseases and their management approaches in India,” in *Sugar beet cultivation, management and processing*. Editors V. Misra, S. Srivastava, and A. K. Mall (Springer Nature Singapore Pte Ltd), 541–564. doi:10.1007/978-981-19-2730-0_26

Mo, W., Tang, W., Du, Y., Jing, Y., Bu, Q., and Lin, R. (2020). Phytochrome interacting Factor-Like14 and Slender Rice1 interaction controls seedling growth under salt stress. *Plant Physiol.* 184, 506–517. doi:10.1104/PP.20.00024

Molteni, V. M., Paris, R., Onofri, C., Orrù, L., Cattivelli, L., Pacifico, D., et al. (2015). Early transcriptional changes in *Beta vulgaris* in response to low temperature. *Planta* 242, 187–201. doi:10.1007/s00425-015-2299-z

Monteiro, F., Frese, L., Castro, S., Duarte, M. C., Paulo, O. S., Loureiro, J., et al. (2018). Genetic and genomic tools to assist sugar beet improvement: the value of the crop wild relatives. *Front. Plant Sci.* 9, 74. doi:10.3389/fpls.2018.00074

Moore, C. E., Meacham-Hensold, K., Lemonnier, P., Slattery, R. A., Benjamin, C., Bernacchi, C. J., et al. (2021). The effect of increasing temperature on crop photosynthesis: from enzymes to ecosystems. *J. Exp. Bot.* 72 (8), 2822–2844. doi:10.1093/jxb/erab090

Mulet, J. M. (2022). “Shaping the sugar beet of tomorrow: current advances in Sugar beet biotechnology and new breeding techniques,” in *Sugar beet cultivation, management and processing*. Editors V. Misra, S. Srivastava, and A. K. Mall (Springer Nature Singapore Pte Ltd), 49–74. doi:10.1007/978-981-19-2730-0_4

Murakami, Y., Tsuyama, M., Kobayashi, Y., Kodama, H., and Iba, K. (2000). Trienoic fatty acids and plant tolerance of high temperature. *Sci* 287, 476–479. doi:10.1126/science.287.5452.476

Mutasa-Gottgens, E. S., Chwarszczynska, D. M., Halsey, K., and Asher, M. J. C. (2000). Specific polyclonal antibodies for the obligate plant parasite *Polymyxa*—a targeted recombinant DNA approach. *Plant Pathol.* 49, 276–287. doi:10.1046/j.1365-3059.2000.00446.x

Nandy, S., Pathak, B., Zhao, S., and Srivastava, V. (2019). Heat-shock-inducible CRISPR/Cas9 system generates heritable mutations in rice. *Plant Direct* 3 (5), e00145. doi:10.1002/pld3.145

Nawaz, G., Han, Y., Usman, B., Liu, F., Qin, B., and Li, R. (2019). Knockout of OsPRP1, a gene encoding proline-rich protein, confers enhanced cold sensitivity in rice (*Oryza sativa* L.) at the seedling stage. *3 Biotech.* 9 (7), 254. doi:10.1007/s13205-019-1787-4

Neher, O. T., and Gallian, J. J. (2013). *Powdery mildew on sugar beet: importance, identification and control*, 643. A Pacific Northwest Extension Publication, 1–6. PNW.

Nielsen, K. K., Bojsen, K., Roepstorff, P., and Mikkelsen, J. D. (1994). A hydroxyproline-containing class IV chitinase of sugar beet is glycosylated with xylose. *Plant Mole. Biol.* 25 (2), 241–257. doi:10.1007/bf00023241

- Nielsen, K. K., Mikkelsen, J. D., Kragh, K. M., and Bojsen, K. (1993). An acidic class III chitinase in sugar beet: induction by *Cercospora beticola*, characterization, and expression in transgenic tobacco plants. *Mol. Plant-Microbe Interact.* 6 (4), 495–506. doi:10.1094/mpmi-6-495
- Nieves-Cordones, M., Mohamed, S., Tanoi, K., Kobayashi, N. I., Takagi, K., Vernet, A., et al. (2017). Production of low-Cs⁺ rice plants by inactivation of the K⁺ transporter of HAK 1 with the CRISPR-Cas system. *Plant J.* 92 (1), 43–56. doi:10.1111/tpj.13632
- Ober, E. S., and Rajabi, A. (2010). Abiotic stress in sugar beet. *Sugar Tech.* 12 (3–4), 294–298. doi:10.1007/s12355-010-0035-3
- Ogata, T., Ishizaki, T., Fujita, M., and Fujita, Y. (2020). CRISPR/Cas9-targeted mutagenesis of *OsERA1* confers enhanced responses to abscisic acid and drought stress and increased primary root growth under non-stressed conditions in rice. *PLoS One* 15, e0243376. doi:10.1371/journal.pone.0243376
- Ordaz, S., Diaz, T., Restrepo, N., Patiño, M. M., and Tamayo, M. C. (1996). Biochemical, immunological and toxicological characteristics of the crystal proteins of *Bacillus thuringiensis* sub sp. modelling. *Mem. Inst. Oswaldo Cruz* 91, 231–237. doi:10.1590/s0074-02761996000200020
- Osakabe, Y., Watanabe, T., Sugano, S. S., Ueta, R., Ishihara, R., Shinokaki, K., et al. (2016). Optimization of CRISPR/Cas9 genome editing to modify abiotic stress responses in plants. *Sci. Rep.* 6 (1), 26685. doi:10.1038/srep26685
- Owen, F. W., and Ryser, G. K. (1942). Some Mendelian characters in *Beta vulgaris* L. and linkages observed in the Y-R-B group. *J. Agric. Res.* 65, 135–171.
- Pankaj, Y. K., and Kumar, V. (2023). “CRISPR/CAS: the Beginning of a new era in crop improvement,”. *Advanced crop improvement*. Editors A. Raina, M. R. Wani, R. A. Laskar, N. Tomlekova, and S. Khan (Cham: Springer), 1, 489–505. doi:10.1007/978-3-031-28146-4_17
- Pardee, K., Green, A. A., Takahashi, M. K., Braff, D., Lambert, G., Lee, J. W., et al. (2016). Rapid, low-cost detection of Zika virus using programmable biomolecular components. *Cell* 165 (5), 1255–1266. doi:10.1016/j.cell.2016.04.059
- Park, C. Y., Lee, J. H., Yoo, J. H., Moon, B. C., Choi, M. S., Kang, Y. H., et al. (2005). WRKY group IId transcription factors interact with calmodulin. *FEBS Lett.* 579, 1545–1550. doi:10.1016/j.febslet.2005.01.057
- Park, J. J., Dempewolf, E., Zhang, W., and Wang, Z. Y. (2017). RNA-guided transcriptional activation via CRISPR/dCas9 mimics overexpression phenotypes in *Arabidopsis*. *PLoS One* 12, e0179410. doi:10.1371/journal.pone.0179410
- Pattanayak, S., Das, S., and Kumar, S. (2023). Development of stress tolerant transgenic traits in sugar beet through biotechnological application. *J. Plant Prot. Res.* 63 (1), 1–12. doi:10.24425/jppr.2023.144505
- Pauchet, Y., Bretschneider, A., Augustin, S., and Heckel, D. G. (2016). A P-glycoprotein is linked to resistance to the *Bacillus thuringiensis* Cry3Aa toxin in a leaf beetle. *Toxins (Basel)* 8, 362. doi:10.3390/toxins8120362
- Paul, H. B., Henken, B., Scholten, O. E., and Lange, W. (1993). Use of zoospores of *Polymyxa betae* in screening beet seedlings for resistance to beet necrotic yellow vein virus. *Neth. J. Plant Pathol.* 99 (3), 151–160. doi:10.1007/bf03041405
- Paul, N. C., Park, S. W., Liu, H., Choi, S., Ma, J., MacCready, J. S., et al. (2021). Plant and fungal genome editing to enhance plant disease resistance using the CRISPR/Cas 9 system. *Front. Plant Sci.* 12, 700925. doi:10.3389/fpls.2021.700925
- Pferdmenges, F. (2007). *Dissertation beet necrotic yellow vein virus. Dissertation*, 23. Cuvillier Verlag Göttingen.
- Piatek, A., Ali, Z., Baazim, H., Li, L., Abulfaraj, A., Al-Shareef, S., et al. (2015). RNA-guided transcriptional regulation in planta via synthetic dCas9-based transcription factors. *Plant Biotechnol. J.* 13 (4), 578–589. doi:10.1111/pbi.12284
- Pidgeon, J. D., Werker, A. R., Jaggard, K. W., Richter, G. M., Lister, D. H., and Jones, P. D. (2001). Climatic impact on the productivity of sugar beet in Europe, 1961–1995. *Agric. Meteorol.* 109, 27–37. doi:10.1016/S0168-1923(01)00254-4
- Porcel, R., Bustamante, A., Ros, R., Serrano, R., and Mulet Salort, J. M. (2018). BvCOLD1: a novel aquaporin from sugar beet (*Beta vulgaris* L.) involved in boron homeostasis and abiotic stress. *Plant Cell Environ.* 41, 2844–2857. doi:10.1111/pce.13416
- Pylypenko, L. A., and Kalatur, K. A. (2015). Breeding and usage of sugar beet cultivars and hybrids resistant to sugar beet nematode *Heterodera schachtii*. *Agric. Sci. Pract.* 2, 12–22. doi:10.15407/agrisp.2.01.012
- Qi, X., Tai, C. Y., and Wasserman, B. P. (1995). Plasma membrane intrinsic proteins of *Beta vulgaris* L. *Plant Physiol.* 108 (1), 387–392. doi:10.1104/pp.108.1.387
- Qiu, Z., Kang, S., He, L., Zhao, J., Zhang, S., Hu, J., et al. (2018). The newly identified heat-stress sensitive albino 1 gene affects chloroplast development in rice. *Plant Sci.* 267, 168–179. doi:10.1016/j.plantsci.2017.11.015
- Rai, K. M., Ghose, K., Rai, A., Singh, H., Srivastava, R., and Mendu, V. (2019). “Genome engineering tools in plant synthetic biology,” in *Current developments in biotechnology and bioengineering*. Editors S. P. Singh, A. Pandey, G. Du, and S. Kumar (Elsevier), 47.
- Ramachandran, V., Weiland, J. J., and Bolton, M. D. (2021). CRISPR-based isothermal next-generation diagnostic method for virus detection in sugarbeet. *Front. Microbio.* 12, 679994. doi:10.3389/fmicb.2021.679994
- Razaq, M. K., Aleem, M., Mansoor, S., Khan, M. A., Rauf, S., Iqbal, S., et al. (2021). Omics and CRISPR-Cas9 approaches for molecular insight, functional gene analysis, and stress tolerance development in crops. *Int. J. Mol. Sci.* 22 (3), 1292. doi:10.3390/ijms22031292
- Regev, A., Keller, M., Strizhov, N., Sneh, B., Prudovsky, E., Chet, I., et al. (1996). Synergistic activity of a *Bacillus thuringiensis* d-endotoxin and a bacterial endochitinase against *Spodoptera littoralis* larvae. *Appl. Environ. Microbiol.* 62, 3581–3586. doi:10.1128/aem.62.10.3581-3586.1996
- Reyer, A., Bazihizina, N., Scherzer, S., Jaslan, J., Schafer, N., Jaslan, D., et al. (2021). Sugar beet cold induced PMT5a and STP 13 carriers are poised for taproot proton-driven plasma membrane sucrose and glucose import. *bioRxiv*. doi:10.1101/2021.09.21.461191
- Roca Paixão, J. F., Gillet, F. X., Ribeiro, T. P., Bournaud, C., Lourenco-Tessutti, I. T., Noriega, D. D., et al. (2019). Improved drought stress tolerance in *Arabidopsis* by CRISPR/dCas9 fusion with a histone Acetyltransferase. *Sci. Rep.* 9, 8080. doi:10.1038/s41598-019-44571-y
- Santeshwari, Misra, V., Mall, A. K., and Kumar, D. (2020). Problems and integrated pest management strategy for *Spodoptera litura* in sugar beet in India. *J. Exp. Zool.* 23 (2), 1887–1890.
- Santosh Kumar, V., Verma, R. K., Yadav, S. K., Yadav, P., Watts, A., Rao, M. V., et al. (2020). CRISPR-Cas9 mediated genome editing of drought and salt tolerance (OsDST) gene in indica mega rice cultivar MTU1010. *Physiol. Mol. Biol. Plants* 26 (6), 1099–1110. doi:10.1007/s12298-020-00819-w
- Schiemann, J., Robiński, J., Schleissig, S., Spok, A., Sprink, T., and Wilhelm, R. A. (2020). Editorial: plant genome editing - policies and governance. *Front. Plant Sci.* 11, 284. doi:10.3389/fpls.2020.00284
- Schmidt, S. M., Belisle, M., and Frommer, W. B. (2020). The evolving landscape around genome editing in agriculture: many countries have exempted or move to exempt forms of genome editing from GMO regulation of crop plants. *EMBO Rep.* 21, e50680. doi:10.15252/embr.202050680
- Scholten, O. E., De Bock, T. S. M., Klein-Lankhorst, R. M., and Lange, W. (1999). Inheritance of resistance to beet necrotic yellow vein virus in *Beta vulgaris* conferred by a second gene for resistance. *Theor. Appl. Genet.* 99, 740–746. doi:10.1007/s001220051292
- Scholten, O. E., Jansen, R. C., Keizer, L. C. P., De Bock, Th. S. M., and Lange, W. (1996). Major genes for resistance to beet necrotic yellow vein virus (BNYVV) in *Beta vulgaris*. *Euphytica* 91, 331–339. doi:10.1007/BF00033095
- Scholten, O. E., Klein-Lankhorst, R. M., Esselink, D. G., De Bock, T. S. M., and Lange, W. (1997). Identification and mapping of random amplified polymorphic DNA (RAPD) markers linked to resistance against beet necrotic yellow vein virus (BNYVV) in *Beta* accessions. *Theor. Appl. Genet.* 94, 123–130. doi:10.1007/s001220050390
- Sedighi, L., Rezapannah, M., and Aghdam, H. R. (2011). Efficacy of Bt transgenic sugar beet lines expressing *cry1Ab* gene against *Spodoptera littoralis* Boisid (Lepidoptera: Noctuidae, Noctuidae). *J. Entomol. Res. Soc.* 13 (1), 61–69.
- Semenova, E., Jore, M. M., Datsenko, K. A., Semenova, A., Westra, E. R., Wanner, B., et al. (2011). Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *PNAS* 108 (25), 10098–10103. doi:10.1073/pnas.1104144108
- Shabbir, R., Singhal, R. K., Mishra, U. N., Chauhan, J., Javed, T., Hussain, S., et al. (2022). Combined abiotic stresses: challenges and potential for crop improvement. *Agronomy* 12 (11), 2795. doi:10.3390/agronomy12112795
- Shah, S. A., Erdmann, S., Mojica, F. J., and Garrett, R. A. (2013). Protospacer recognition motifs: mixed identities and functional diversity. *RNA Biol.* 10 (5), 891–899. doi:10.4161/rna.23764
- Sharma, P., and Dubey, R. S. (2005). Lead toxicity in plants. *Braz. J. Plant Physiol.* 17, 35–52. doi:10.1590/S1677-04202005000100004
- Shen, C., Que, Z., Xia, Y., Tang, N., Li, D., He, R., et al. (2017). Knock out of the annexin gene OsAnn3 via CRISPR/Cas9-mediated genome editing decreased cold tolerance in rice. *J. Plant Biol.* 60, 539–547. doi:10.1007/s12374-016-0400-1
- Shivakumara, T. N., Somvanshi, V. S., Phani, V., Chaudhary, S., Hada, A., Budhwar, R., et al. (2019). *Meloidogyne incognita* (Nematoda: meloidogynidae) sterol-binding protein Mi-SBP-1 as a target for its management. *Int. J. Parasitol.* 49 (13–14), 1061–1073. doi:10.1016/j.ijpara.2019.09.002
- Singh, S., Kumar, N. R., Maniraj, R., Lakshminanth, R., Rao, K. Y. S., Muralimohan, N., et al. (2018). Expression of Cry2Aa, a *Bacillus thuringiensis* insecticidal protein in transgenic pigeon pea confers resistance to Gram pod borer, *Helicoverpa armigera*. *Sci. Rep.* 8, 8820. doi:10.1038/s41598-018-26358-9
- Singh, S., Parihar, P., Singh, R., Singh, V. P., and Prasad, S. M. (2016). Heavy metal tolerance in plants: role of transcriptomics, proteomics, metabolomics and ionomics. *Front. Plant Sci.* 6, 1143. doi:10.3389/fpls.2015.01143
- Smigocki, A. C., Ivic, S. D., Wilson, D., Wozniak, C. A., Campbell, L., Dregseth, R., et al. (2003). “Molecular approaches for control of the sugar beet root maggot,” in 1st joint IIRB-ASSBT Congress, San Antonio, USA, 26th Feb-1st March 2003, 416–428.
- Smith, G. A., and Ruppel, E. G. (1973). Association of *Cercospora* leaf spot, gross sucrose, percentage sucrose, and root weight in sugar beet. *Can. J. Plant Sci.* 53, 695–696. doi:10.4141/cjps73-136

- Stevanato, P., Chiodi, C., Broccanello, C., Concheri, G., Biancardi, E., Pavli, O., et al. (2019). Sustainability of the sugar beet crop. *Sugar Tech.* 21, 703–716. doi:10.1007/s12355-019-00734-9
- Stewart, S. D., Adamczyk, J. J., Kinghten, K. S., and Davis, F. M. (2001). Impact of *Bt* cotton expressing one or two insecticidal proteins of *Bacillus thuringiensis* Berliner on growth and survival of Noctuidae (Lepidoptera) larvae. *J. Econ. Entomol.* 94 (3), 752–760. doi:10.1603/0022-0493.94.3.752
- Stokstad, E. (2021). U.K. set to loosen rules for gene-edited crops and animals. *Sci. News Sci. Insid.* doi:10.1126/science.abj6955
- Strausbaugh, C. A., Eujayl, I. A., and Wintermantel, W. M. (2017). Beet curly top virus strains associated with sugar beet in Idaho, Oregon, and a Western U.S. Collection. *Plant Dis.* 101 (8), 1373–1382. doi:10.1094/PDIS-03-17-0381-RE
- Strausbaugh, C. A., Wenninger, E. J., and Eujayl, I. A. (2012). Management of severe curly top in sugar beet with insecticides. *Plant Dis.* 96, 1159–1164. doi:10.1094/PDIS-01-12-0106-RE
- Sun, D., Guo, Z., Liu, Y., and Zhang, Y. (2017). Progress and prospects of CRISPR/Cas systems in insects and other arthropods. *Front. Physiol.* 8, 608. doi:10.3389/fphys.2017.00608
- Sun, J., Hu, W., Zhou, R., Wang, L., Wang, X., Wang, Q., et al. (2014). The *Brachypodium distachyon* BdWRKY36 gene confers tolerance to drought stress in transgenic tobacco plants. *Plant Cell Rep.* 34, 23–35. doi:10.1007/s00299-014-1684-6
- Taguchi, K., Kubo, T., Takahashi, H., and Abe, H. (2011). Identification and precise mapping of resistant QTLs of *Cercospora* leaf spot resistance in sugar beet (*Beta vulgaris* L. G3 (Bethesda) 1 (4), 283–291. doi:10.1534/g3.111.000513
- Taguchi, K., Okazaki, K., Takashi, H., Kubo, T., and Mikami, T. (2010). Molecular mapping of a gene conferring resistance to *Aphanomyces* root rot (black rot) in sugar beet (*Beta vulgaris* L.). *Euphytica* 173, 408–418. doi:10.1007/s10681-010-0153-8
- Takenaka, S., and Tamagake, H. (2009). Foliar spray of a cell wall protein fraction from the biocontrol agent *Pythium oligandrum* induces defence-related genes and increases resistance against *Cercospora* leaf spot in sugar beet. *J. Gen. Plant Pathol.* 75 (5), 340–348. doi:10.1007/s10327-009-0186-9
- Tan, W., Li, K., Liu, D., and Xing, W. (2023). *Cercospora* leaf spot disease of sugar beet. *Plant Signal Behav.* 18 (1), 2214765. doi:10.1080/15592324.2023.2214765
- Tang, X., Lowder, L. G., Zhang, T., Malzahn, A. A., Zheng, X., Voytas, D. F., et al. (2017). Correction: a CRISPR–Cpf1 system for efficient genome editing and transcriptional repression in plants. *Nat. Plants* 3 (7), 17103. doi:10.1038/nplants.2017.103
- Thakur, M., Praveen, S., Divte, P. R., Mitra, R., Kumar, M., Gupta, C. K., et al. (2022). Metal tolerance in plants: molecular and physicochemical interface determines the “not so heavy effect” of heavy metals. *Chemosphere* 287 (1), 131957. doi:10.1016/j.chemosphere.2021.131957
- Thomas, M., Parry-Smith, D., and Iyer, V. (2019). Best practice for CRISPR design using current tools and resources. *Methods* 164, 3–17. doi:10.1016/j.ymeth.2019.05.019
- Thurau, T., Kifle, S., Jung, C., and Cai, D. (2003). The promoter of the nematode resistance gene *Hs1^{PMO-1}* activates a nematode responsive and feeding site-specific gene expression in sugar beet (*Beta vulgaris* L.) and *Arabidopsis thaliana*. *Plant Mol. Biol.* 52 (3), 643–660. doi:10.1023/a:1024887516581
- Trela, Z., Burdach, Z., Przestalski, S., and Karcz, W. (2012). Effect of trimethyllead chloride on slowly activating (SV) channels in red beet (*Beta vulgaris* L.) taproots. *Comptes Rendus Biol.* 335, 722–730. doi:10.1016/j.crv.2012.11.004
- Uphoff, H., and Wricke, G. (1992). Random amplified polymorphic DNA (RAPD) markers in sugar beet (*Beta vulgaris* L.): mapping the genes for nematode resistance and hypocotyl colour. *Plant Breed.* 109, 168–171. doi:10.1111/j.1439-0523.1992.tb00167.x
- Vinson, C. C., Mota, A. P. Z., Porto, B. N., Oliveira, T. N., Sampaio, I., Lacerda, A. L., et al. (2020). Characterization of raffinose metabolism genes uncovers a wild *Arachis* galactinol synthase conferring tolerance to abiotic stresses. *Sci. Rep.* 10, 15258. doi:10.1038/s41598-020-72191-4
- Wan, L., Wang, Z., Tang, M., Hong, D., Sun, Y., Ren, J., et al. (2021). CRISPR-Cas9 gene editing for fruit and vegetable crops: strategies and prospects. *Hortic* 7, 193. doi:10.3390/horticulturae7070193
- Wang, B., Xie, G., Liu, Z., He, R., Han, J., Huang, S., et al. (2019). Mutagenesis reveals that the OsPPa6 gene is required for enhancing the alkaline tolerance in rice. *Front. Plant Sci.* 10, 759. doi:10.3389/fpls.2019.00759
- Wang, B., Zhong, Z., Wang, X., Han, X., Yu, D., Wang, C., et al. (2020). Knockout of the OsNAC006 transcription factor causes drought and heat sensitivity in rice. *Int. J. Mol. Sci.* 21, 2288. doi:10.3390/ijms21072288
- Wang, F., Hou, X., Tang, J., Wang, Z., Wang, S., Jiang, F., et al. (2011). A novel cold-inducible gene from Pak-choi (*Brassica campestris* ssp. *chinensis*), BcWRKY46, enhances the cold, salt and dehydration stress tolerance in transgenic tobacco. *Mol. Biol. Rep.* 39, 4553–4564. doi:10.1007/s11033-011-1245-9
- Wang, F. Z., Chen, M. X., Yu, L. J., Xie, L. J., Yuan, L.-B., Qi, H., et al. (2017b). OsARM1, an R2R3 MYB transcription factor, is involved in regulation of the response to arsenic stress in rice. *Front. Plant Sci.* 8, 1868. doi:10.3389/fpls.2017.01868
- Wang, G., Liang, G., Wu, K., and Guo, Y. (2005). Gene cloning and sequencing of aminopeptidase N3, a putative receptor for *Bacillus thuringiensis* insecticidal CryIAc toxin in *Helicoverpa armigera* (Lepidoptera: noctuidae). *Eur. J. Entomol.* 102, 13–19. doi:10.14411/eje.2005.002
- Wang, H., Shi, Y., Wang, L., Liu, S., Wu, S., Yang, Y., et al. (2018). CYP6AE gene cluster knockout in *Helicoverpa armigera* reveals role in detoxification of phytochemicals and insecticides. *Nat. Commun.* 9, 4820. doi:10.1038/s41467-018-07226-6
- Wang, J., Wang, H., Liu, S., Liu, T., Tay, W. T., Walsh, T. K., et al. (2017a). CRISPR/Cas9 mediated genome editing of *Helicoverpa armigera* with mutations of an ABC transporter gene HaABCA2 confers resistance to *Bacillus thuringiensis* Cry2A toxins. *Insect biochem. Mol. Biol.* 87, 147–153. doi:10.1016/j.ibmb.2017.07.002
- Wang, J., Zhang, H., Wang, H., Zhao, S., Zuo, Y., Yang, Y., et al. (2016). Functional validation of cadherin as a receptor of Bt toxin Cry1Ac in *Helicoverpa armigera* utilizing the CRISPR/Cas9 system. *Insect biochem. Mol. Biol.* 76, 11–17. doi:10.1016/j.ibmb.2016.06.008
- Wang, M., Vannozzi, A., Wang, G., Liang, Y. H., Tornielli, G. B., Zenoni, S., et al. (2014). Genome and transcriptome analysis of the grapevine (*Vitis vinifera* L.) WRKY gene family. *Hortic. Res.* 1, 14016. doi:10.1038/hortres.2014.16
- Wang, S., Yi, F., and Qu, J. (2015a). Eliminate mitochondrial diseases by gene editing in germ-line cells and embryos. *Protein Cell* 6, 472–475. doi:10.1007/s13238-015-0177-x
- Wang, T., Xun, H., Wang, W., Ding, X., Tian, H., Hussain, S., et al. (2021b). Mutation of GmA1TR genes by CRISPR/cas9 genome editing results in enhanced salinity stress tolerance in soybean. *Front. Plant Sci.* 12, 779598. doi:10.3389/fpls.2021.779598
- Wang, W., Vinocur, B., and Altman, A. (2003). Plant responses to drought, salinity and extreme temperatures: towards genetic engineering for stress tolerance. *Planta* 218, 1–14. doi:10.1007/s00425-003-1105-5
- Wang, Y., Stevanato, P., Yu, L., Zhao, H., Sun, X., Sun, F., et al. (2017). The physiological and metabolic changes in sugar beet seedlings under different levels of salt stress. *J. Plant Res.* 130, 1079–1093. doi:10.1007/s10265-017-0964-y
- Wang, Y., Wang, S., Tian, Y., Wang, Q., Chen, S., Li, H., et al. (2021a). Functional characterization of a sugar beet BVBHLH93 transcription factor in salt stress tolerance. *Int. J. Mol. Sci.* 22, 3669. doi:10.3390/ijms22073669
- Wang, Y. H., Chen, X. E., Yang, Y., Xu, J., Fang, G. Q., Niu, C. Y., et al. (2019). The Masc gene product controls masculinization in the black cutworm, *Agrotis ipsilon*. *Insect Sci.* 26 (6), 1037–1044. doi:10.1111/1744-7917.12635
- Wang, Z., Wang, Y., Tong, Q., Xu, G., Xu, M., Li, H., et al. (2021). Transcriptomic analysis of grapevine Dof transcription factor gene family in response to cold stress and functional analyses of the VaDof17d gene. *Planta* 253, 55. doi:10.1007/s00425-021-03574-8
- War, A. R., Paulraj, M. G., Ahmad, T., Buhroo, A. A., Hussain, B., Ignacimuthu, S., et al. (2012). Mechanisms of plant defense against insect herbivores. *Plant Signal Behav.* 7, 1306–1320. doi:10.4161/psb.21663
- Westra, E. R., Semenova, E., Datsenko, K. A., Jackson, R. N., Wiedenheft, B., Severinov, K., et al. (2013). Type IE CRISPR-cas systems discriminate target from non-target DNA through base pairing-independent PAM recognition. *PLoS Genet.* 9 (9), e1003742. doi:10.1371/journal.pgen.1003742
- Wetzel, V., Willems, G., DarracqGalein, Y., Liebe, S., Varrelmann, T., and Varrelmann, M. (2021). The *Beta vulgaris* derived resistance gene Rz2 confers broad spectrum resistance against soil borne sugar beet infecting viruses from different families by recognizing triple gene block protein 1. *Mol. Plant Pathol.* 22 (7), 829–842. doi:10.1111/mpp.13066
- Wilhite, S. E., Elden, T. C., Puizdar, V., Armstrong, S., and Smigocki, A. C. (2000). Inhibition of aspartyl and serine proteinases in the midgut of sugarbeet root maggot with proteinase inhibitors. *Entomol. Exp. Appl.* 97, 229–233. doi:10.1046/j.1570-7458.2000.00734.x
- Windsels, C. E. (2000). *Aphanomyces* root rot on sugar beet. *Plant Health Prog.* 1 (1), 1–6. doi:10.1094/PHP-2000-0720-01-DG
- Wu, G. Q., Li, Z. Q., Cao, H., and Wang, J. L. (2019). Genome-wide identification and expression analysis of the WRKY genes in sugar beet (*Beta vulgaris* L.) under alkaline stress. *Peer J.* 7, e7817. doi:10.7717/peerj.7817
- Wu, J., Yan, G., Duan, Z., Wang, Z., Kang, C., Guo, L., et al. (2020). Roles of the *Brassica napus* DELLA protein BnaA6. RGA, in modulating drought tolerance by interacting with the ABA signalling component BnaA10. *ABF2*. *Front. Plant Sci.* 11, 577. doi:10.3389/fpls.2020.00577
- Wu, T. Y., Hoh, K. L., Boonyaves, K., Krishnamoorthi, S., and Urano, D. (2022). Diversification of heat shock transcription factors expanded thermal stress responses during early plant evolution. *Plant Cell* 34 (10), 3557–3576. doi:10.1093/plcell/koac204
- Yang, Y., Wang, Y. H., Chen, X. E., Tian, D., Xu, X., Li, K., et al. (2018). CRISPR/Cas9-mediated Tyrosine hydroxylase knockout resulting in larval lethality in *Agrotis ipsilon*. *Insect Sci.* 25 (6), 1017–1024. doi:10.1111/1744-7917.12647
- Yerzhebayeva, R., Abekova, A., Konysebekov, K., Bastaubayeva, S., Kabdrakhmanova, A., Absattarova, A., et al. (2018). Two sugar beet chitinase genes, BvSP2 and BvSE2, analysed with SNP Amplifluor-like markers, are highly expressed after Fusarium root rot inoculations and field susceptibility trial. *Peer J.* 6, e5127. doi:10.7717/peerj.5127
- Yildirim, K., Kavas, M., Kucuk, İ. S., Secgin, Z., and Sarac, C. G. (2023). Development of highly efficient resistance to beet curly top Iran virus (Becurtovirus) in sugar beet (*B. vulgaris*) via CRISPR/Cas9 System. *Int. J. Mol. Sci.* 24, 6515. doi:10.3390/ijms24076515

- Yildirim, K., Secgin, Z., Senyer, A., Can, C., and Kavas, M. (2019). "Conferring multiple resistance to DNA viruses in plants with CRISPR/Cas9 genome editing technology," in 1st Plant Ed Conference Plant genome editing at State of the Art, Novi Sad.
- Yildirim, K., Kavas, M., Kaya, R., Secgin, Z., Can, C., Sevgen, I., et al. (2022). Genome based identification of beet curly top Iran virus infecting sugar beet in Turkey and investigation of its pathogenicity by agroinfection. *J. Virol. Methods* 300, 114380. doi:10.1016/j.jviromet.2021.114380
- Yin, K., Gao, C., and Qiu, J. L. (2017). Progress and prospects in plant genome editing. *Nat. Plants* 3, 17107. doi:10.1038/nplants.2017.107
- Yin, W., Xiao, Y., Niu, M., Meng, W., Li, L., Zhang, X., et al. (2020). ARGONAUTE2 enhances grain length and salt tolerance by activating BIG GRAIN3 to modulate cytokinin distribution in rice. *Plant Cell* 32, 2292–2306. doi:10.1105/tpc.19.00542
- Yolcu, S., Alavilli, H., Ganesh, P., Asif, M., Kumar, M., and Song, K. (2021). An insight into the abiotic stress responses of cultivated beets (*Beta vulgaris* L.). *Plants* 11 (1), 12. doi:10.3390/plants11010012
- Yu, B., Chen, M., Grin, I., and Ma, C. (2020). "Mechanisms of sugar beet response to biotic and abiotic stresses," in *Mechanisms of genome protection and repair*. Editor D. O. Zharkov (Cham: Springer), 167–194. doi:10.1007/978-3-030-41283-8_10
- Yuan, Q., Metterville, D., Briscoe, A. D., and Reppert, S. M. (2007). Insect cryptochromes: gene duplication and loss define diverse ways to construct insect circadian clocks. *Mol. Biol. Evol.* 24 (4), 948–955. doi:10.1093/molbev/msm011
- Zeng, D. D., Yang, C. C., Qin, R., Alamin, M., Yue, E. K., Jin, X. L., et al. (2018). A guanine insert in OsBBS1 leads to early leaf senescence and salt stress sensitivity in rice (*Oryza sativa* L.). *Plant Cell Rep.* 37, 933–946. doi:10.1007/s00299-018-2280-y
- Zeng, Y., Wen, J., Zhao, W., Wang, Q., and Huang, W. (2020). Rational improvement of rice yield and cold tolerance by editing the three genes OsPIN5b, GS3, and OsMYB30 with the CRISPR–Cas9 system. *Front. Plant Sci.* 10, 1663–1946. doi:10.3389/fpls.2019.01663
- Zetsche, B., Gootenberg, J. S., Abudayyeh, O. O., Slaymaker, I. M., Makarova, K. S., Essletzbichler, P., et al. (2015). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* 163 (3), 759–771. doi:10.1016/j.cell.2015.09.038
- Zhang, A., Liu, Y., Wang, F., Li, T., Chen, Z., Kong, D., et al. (2019). Enhanced rice salinity tolerance via CRISPR/Cas9-targeted mutagenesis of the OsRR22 gene. *Mol. Breed.* 39, 47–10. doi:10.1007/s11032-019-0954-y
- Zhang, C. L., Xu, D. C., Jiang, X. C., Zhou, Y., Cui, J., Zhang, C. X., et al. (2008). Genetic approaches to sustainable pest management in sugar beet (*Beta vulgaris* L.). *Ann. Appl. Biol.* 152 (2), 143–156. doi:10.1111/j.1744-7348.2008.00228.x
- Zhang, J., Jin, M., Yang Yliu, L., Yang, Y., Gómez, I., Bravo, A., et al. (2020a). The cadherin protein is not involved in susceptibility to *Bacillus thuringiensis* Cry1Ab or Cry1Fa toxins in *Spodoptera frugiperda*. *Toxins (Basel)* 12 (6), 375. doi:10.3390/toxins12060375
- Zhang, X., Long, Y., Huang, J., and Xia, J. (2020b). OsNAC45 is involved in ABA response and salt tolerance in rice. *Rice* 13, 79–13. doi:10.1186/s12284-020-00440-1
- Zhang, Y., Guo, W., Chen, L., Shen, X., Yang, H., Fang, Y., et al. (2022). CRISPR/Cas9-mediated targeted mutagenesis of GmUGT enhanced soybean resistance against leaf-chewing insects through flavonoids biosynthesis. *Front. Plant Sci.* 13, 802716. doi:10.3389/fpls.2022.802716
- Zhao, Y., Zhang, C., Liu, W., Gao, W., Liu, C., Song, G., et al. (2016). An alternative strategy for targeted gene replacement in plants using a dual-sgRNA/Cas9 design. *Sci. Rep.* 6, 23890. doi:10.1038/srep23890
- Zheng, J. C., Yue, X. R., Kuang, W. Q., Li, S. L., Tang, R., Zhang, Z. F., et al. (2020). NPC1b as a novel target in controlling the cotton bollworm, *Helicoverpa armigera*. *Pest Manag. Sci.* 76, 2233–2242. doi:10.1002/ps.5761
- Zhou, C., Liu, D., Wu, P., Wang, Y., Gai, Z., Liu, L., et al. (2020). Transcriptome analysis of sugar beet (*Beta vulgaris* L.) in response to alkaline stress. *Plant mole. Biol.* 102, 645–657. doi:10.1007/s11103-020-00971-7
- Zhu, G. H., Cherredy, S. C., Howell, J. L., and Palli, S. R. (2020). Genome editing in the fall armyworm, *Spodoptera frugiperda*: multiple sgRNA/Cas9 method for identification of knockouts in one generation. *Insect biochem. Mol. Biol.* 122, 103373. doi:10.1016/j.ibmb.2020.103373
- Zhu, G. H., Peng, Y. C., Zheng, M. Y., Zhang, X. Q., Sun, J. B., Huang, Y., et al. (2017). CRISPR/Cas9 mediated BLOS2 knockout resulting in disappearance of yellow strips and white spots on the larval integument in *Spodoptera litura*. *J. Insect Physiol.* 103, 29–35. doi:10.1016/j.jinsphys.2017.09.008
- Zhu, G. H., Xu, J., Cui, Z., Dong, X. T., Ye, Z. F., Niu, D. J., et al. (2016). Functional characterization of SlitPBP3 in *Spodoptera litura* by CRISPR/Cas9 mediated genome editing. *Insect biochem. Mol. Biol.* 75, 1–9. doi:10.1016/j.ibmb.2016.05.006
- Zhu, Z., Shi, J., Cao, J., He, M., and Wang, Y. (2012). VpWRKY3, a biotic and abiotic stress-related transcription factor from the Chinese wild *Vitis pseudoreticulata*. *Plant Cell Rep.* 31, 2109–2120. doi:10.1007/s00299-012-1321-1
- Zicari, S., Zhang, R., and Kaffka, S. (2019). "Sugar beet," in *Integrated processing technologies for food and agricultural by-products*. Editors Z. Pan, R. Zhang, and S. Zircari (Apple Academic Press), 331–351. doi:10.1016/b978-0-12-814138-0.00013-7
- Zou, C., Jiang, W., and Yu, D. (2010). Male gametophyte-specific WRKY34 transcription factor mediates cold sensitivity of mature pollen in *Arabidopsis*. *J. Exp. Bot.* 61, 3901–3914. doi:10.1093/jxb/erq204
- Zou, C., Wang, Y., Wang, B., Liu, D., Liu, L., Gai, Z., et al. (2020). Long non-coding RNAs in the alkaline stress response in sugar beet (*Beta vulgaris* L.). *BMC Plant Biol.* 20, 227. doi:10.1186/s12870-020-02437-w
- Zuo, Y., Shi, Y., Zhang, F., Guan, F., Zhang, J., Feyereisen, R., et al. (2021). Genome mapping coupled with CRISPR gene editing reveals a P450 gene confers avermectin resistance in the beet armyworm. *PLoS Genet.* 17 (7), e1009680. doi:10.1371/journal.pgen.1009680
- Zuo, Y., Wang, H., Xu, Y., Huang, J., Wu, S., Wu, Y., et al. (2017). CRISPR/Cas9 mediated G4946E substitution in the ryanodine receptor of *Spodoptera exigua* confers high levels of resistance to diamide insecticides. *Insect biochem. Mol. Biol.* 89, 79–85. doi:10.1016/j.ibmb.2017.09.005
- Zuo, Y., Xue, Y., Lu, W., Ma, H., Chen, M., Wu, Y., et al. (2020). Functional validation of nicotinic acetylcholine receptor (nAChR) $\alpha 6$ as a target of spinosyns in *Spodoptera exigua* utilizing the CRISPR/Cas9 system. *Pest Manag. Sci.* 76, 2415–2422. doi:10.1002/ps.5782
- Zuo, Y. Y., Huang, J. L., Wang, J., Feng, Y., Han, T. T., Wu, Y. D., et al. (2018). Knockout of a P-glycoprotein gene increases susceptibility to abamectin and emamectin benzoate in *Spodoptera exigua*. *Insect Mol. Biol.* 27, 36–45. doi:10.1111/imb.12338

Glossary

aaNAT	Arylalkylamine-N-acetyltransferase	Pt	Platinum
Ag	Silver	RKN	Root-Knot Nematode
APX	Ascorbate Peroxidase	ROS	Reactive oxygen species
As	Arsenic	RPA	Recombinase Polymerase Amplification
BBMV	Brush-border membrane vesicles	SBRM	Sugar beet root maggot
BCTV	Beet Curly Top Virus Resistance	sgRNA	Short Guide RNA
BNYVV	Beet Necrotic Yellow Vein Virus	SOD	Superoxide Dismutase
CAT	Catalase	TALENs	Transcription Activator Like Effector Nucleases
Cd	Cadmium	TSS	Transcription Start Site
Co	Cobalt	VIP	Vegetal Insecticidal Protein
CP	Capsid Protein	ZFNs	Zinc-Finger Nucleases
Cr	Chromium	Zn	Zinc
CRISPR/Cas	Clustered Regularly Interspaced Short Palindromic Repeats		
CRISPRa	CRISPR Activation		
CRISPRi	CRISPR Interference		
crRNA	CRISPR RNA		
CRY	Crystalline Protein		
DE	Differentially Expressed		
DETECTR	DNA Endonuclease Targeted CRISPR Trans Reporter		
dsRNA	Double-Stranded RNA		
Fe	Iron		
FQ	Fluorescence Quencher		
GR	Glutathione Reductase		
GVR	Geminiviral Replicons		
HD-RNAi	Host-Delivered RNA Interference		
HIGS	Host-Induced Gene Silencing		
HSFs	Heat Shock Transcription Factors		
HSPs	Heat Shock Proteins		
HSEs	Heat Shock Elements		
ICPs	Insecticidal Crystal Proteins		
KRAB	Kruppel Associated Box		
lncRNAs	Long noncoding RNAs		
NADA	N-acetyl dopamine		
Ni	Nickel		
NIC	Toxic Nickel Concentration		
PAM	Protospacer Adjacent Motif		
Pb	Lead		
PI	Protease Inhibitor		
PM	Peritrophic Membrane		
POC	Point-of-Care		



OPEN ACCESS

EDITED BY

Rupesh Deshmukh,
Central University of Haryana, India

REVIEWED BY

Pandiyan Muthuramalingam,
Gyeongsang National University,
Republic of Korea
Himanshu Pandey,
Khalsa College, Amritsar, India

*CORRESPONDENCE

Yahui Chen,
✉ chenياهو@njfu.edu.cn
Jiang Jiang,
✉ ecologyjiang@gmail.com
Lei Wang,
✉ wl.stone@163.com

RECEIVED 08 August 2023

ACCEPTED 30 October 2023

PUBLISHED 20 November 2023

CITATION

Chen Y, Zhang X, Fan Y, Sui D, Jiang J and Wang L (2023), The role of WRKY transcription factors in exogenous potassium (K⁺) response to NaCl stress in *Tamarix ramosissima*. *Front. Genet.* 14:1274288. doi: 10.3389/fgene.2023.1274288

COPYRIGHT

© 2023 Chen, Zhang, Fan, Sui, Jiang and Wang. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). 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.

The role of WRKY transcription factors in exogenous potassium (K⁺) response to NaCl stress in *Tamarix ramosissima*

Yahui Chen^{1,2*}, Xuanyi Zhang^{1,2}, Yunlong Fan³, Dezong Sui¹, Jiang Jiang^{2*} and Lei Wang^{1*}

¹Jiangsu Academy of Forestry, Nanjing, China, ²Collaborative Innovation Center of Sustainable Forestry in Southern China of Jiangsu Province, Nanjing Forestry University, Nanjing, China, ³Faculty of Science Department of Statistics, University of British Columbia, Vancouver, BC, Canada

Introduction: Soil salinization poses a significant challenge to plant growth and vitality. Plants like *Tamarix ramosissima* Ledeb (*T. ramosissima*), which are halophytes, are often integrated into planting schemes tailored for saline environments. Yet, the role of WRKY transcription factors in *T. ramosissima*, especially under sodium chloride (NaCl) stress mitigated by exogenous K⁺ application, is not well-understood. This research endeavors to bridge this knowledge gap.

Methods: Using Pfam protein domain prediction and physicochemical property analysis, we delved into the WRKY genes in *T. ramosissima* roots that are implicated in counteracting NaCl stress when aided by exogenous K⁺ applications. By observing shifts in the expression levels of WRKY genes annotated to the KEGG pathway under NaCl stress at 0, 48, and 168 h, we aimed to identify potential key WRKY genes.

Results: We found that the expression of 56 WRKY genes in *T. ramosissima* roots responded to exogenous K⁺ application during NaCl stress at the indicated time points. Particularly, the expression levels of these genes were primarily upregulated within 168 h. From these, 10 WRKY genes were found to be relevant in the KEGG pathways. Moreover, six genes, namely *Unigene0024962*, *Unigene0024963*, *Unigene0010090*, *Unigene0007135*, *Unigene0070215*, and *Unigene0077293*, were annotated to the Plant-pathogen interaction pathway or the MAPK signaling pathway in plants. These genes exhibited dynamic expression regulation at 48 h with the application of exogenous K⁺ under NaCl stress.

Discussion: Our research highlights that WRKY transcription factors can modulate the activation or inhibition of related genes during NaCl stress with the application of exogenous K⁺. This regulation enhances the plant's adaptability to saline environments and mitigates the damage induced by NaCl. These findings provide valuable gene resources for future salt-tolerant *Tamarix* breeding and expand our understanding of the molecular mechanisms of WRKY transcription factors in alleviating NaCl toxicity.

KEYWORDS

exogenous potassium (K⁺), halophyte, NaCl stress, transcriptome, WRKY transcription factor

1 Introduction

Soil salinization poses a significant environmental issue worldwide. Reports indicate that approximately 8.31×10^6 km² of the earth's soil is affected by salinity (Li et al., 2014), and projections warn that due to escalating soil salinity, as much as 50% of the world's arable land could be lost by mid-century. This exacerbation is predominantly caused by environmental degradation, global warming, and inappropriate irrigation practices (Ludwig et al., 2018). High salt levels are known to inhibit plant growth severely, a phenomenon first highlighted in Boyer's 1984 study (Boyer, 1982). The ramifications on global agriculture are considerable, with an estimated economic toll exceeding 27 billion US dollars annually due to loss in productivity (Qadir et al., 2014). Therefore, devising strategies to utilize saline-alkaline land efficiently and promoting the development and cultivation of salt-resistant crops is of paramount importance (Savary et al., 2020).

Salt stress, the detrimental impact of highly saline soils on plant growth, affects many stages of a plant's life, from seed germination to flowering and fruiting (Park et al., 2013). This often results in reduced germination rates, slower growth, decreased plant height, and leaf wilting (Hao et al., 2021). High Na⁺ concentrations in saline-alkali soils cause osmotic stress, obstructing the absorption of water and nutrients, leading to a nutrient-water imbalance (Mauro et al., 2022). Potassium (K⁺) is a key cation in plant cells, improving plant resilience to both abiotic stresses, such as salt, drought, and heavy metals, and biotic stresses like fungi (Leigh and Jones, 1984; Amtmann et al., 2008). High salt content in soil hinders K⁺ absorption due to competition with Na⁺, leading to a decrease in the K⁺/Na⁺ ratio, which results in excessive Na⁺ in plants, inhibiting growth and potentially causing plant death (Shabala and Cuin, 2008; Wang et al., 2019). Plant enrichment with K⁺ can enhance their water status, biomass, and salt tolerance (Tittal et al., 2021; Johnson et al., 2022). Under salt stress, the Na⁺ content in plants greatly exceeds K⁺ levels (Chakraborty et al., 2016). To prevent Na⁺ toxicity, plants must maintain a high K⁺/Na⁺ ratio to regulate the Na⁺/K⁺ balance (Chakraborty et al., 2016). Thus, boosting K⁺ absorption and minimizing Na⁺ accumulation in plants is a prime strategy against salt stress (Munns and Tester, 2008; Shabala and Cuin, 2008; Chen et al., 2022a).

Transcription factors (TFs) are also known as trans-acting factors and are specialized proteins that can regulate plant growth and development. Transcription factors can bind to specific DNA sequences (cis-acting elements) in the promoter region of target genes through their DNA-binding domains (DBD). This interaction allows them to control the specific expression of target genes in different tissues, cells, or under various environmental conditions in plants (Singh et al., 2002). Transcription factor families such as WRKY, V-myb avian myeloblastosis viral oncogene homolog (MYB), NAM, ATAF1/2 and CUC2 (NAC), Basic leucine zipper (bZIP), AP2/ERF, and Basic helix-loop-helix (bHLH) play pivotal regulatory and molecular switch roles in the intricate salt stress signaling network. By activating or suppressing the expression of specific genes or a set of genes, these transcription factors ensure their targeted expression. The products resulting from this gene expression further control the expression of downstream

genes or directly act to protect plants from the damage induced by salt stress (Khan et al., 2018). It's worth noting that WRKY transcription factors have been recognized over the past two decades as instrumental in regulating plant responses to biotic and abiotic stresses (Schlüttenhofer and Yuan, 2015).

WRKY transcription factors are a plant-specific transcription factor family, which is one of the largest transcription factor families in higher plants (Ulker and Somssich, 2004; Cheng et al., 2021). These factors regulate gene transcription by identifying and binding to the W-box, a conserved DNA binding site. As a result, they play roles in various plant functions, including growth, development, defense against pathogens, responses to both biological and environmental stress, hormone signal processing, secondary metabolism, and other processes (Li and Zhou, 2014; Wei et al., 2022). WRKY transcription factors can also regulate the expression of related genes by binding cis-elements or interacting with other regulatory factors to play an important role in plant growth and development and resist abiotic stress (Wei et al., 2022). In particular, WRKY transcription factors can positively or negatively regulate the expression of related genes when plants are subjected to salt stress, thus making plants tolerant to salt. Dai et al. reported that *Fortunella crassifolia* FcWRKY40 can directly activate the expression of serine/threonine protein kinase gene FcSOS2 in SOS pathway, indirectly regulate the expression of FcSOS1 and FcSOS3 genes, promote Na⁺ effection, and positive response to salt stress (Dai et al., 2018). Research by Song and colleagues revealed that SbWRKY50, found in Sorghum, had a negative impact on the salt response by down-regulating the expression of AtSOS1, a Na⁺/H⁺ reverse transporter gene in *Arabidopsis* (Song et al., 2020). The overexpression of OsWRKY40 in rice decreased the tolerance of rice to high salt (Tao et al., 2011). GhWRKY25 is overexpressed and its tolerance to salt stress is enhanced, while in transgenic tobacco, it has relatively weak tolerance to drought stress (Liu et al., 2016).

As halophytes, *Tamarix* plants utilize their innate physiological and metabolic mechanisms to enhance their resistance to saline conditions (Gao et al., 2011; Gupta et al., 2014; Che et al., 2019; Bahramsoltani et al., 2020). Plants of the *Tamarix* plants have remarkable resilience to high salt environments, making them essential for restoring arid, desert, and saline regions (Yang et al., 2021; Sun et al., 2022). Recognizing their value, China has significantly invested in *Tamarix* cultivation as part of its ecological conservation initiatives. Additionally, *Tamarix* plants play a vital role in Chinese herbal medicine due to their medicinal properties (Duan et al., 2022). Studies by Flowers and Colmer indicate that halophytes can successfully complete their life cycle even when subjected to NaCl concentrations exceeding 200 mM (Flowers and Colmer, 2008; Liu et al., 2022). Typically, *Tamarix ramosissima* Ledeb (*T. ramosissima*), a representative of the *Tamarix* plants, is used as a restoration species in saline-alkali areas (Duan et al., 2022). *T. ramosissima* can sustain growth under NaCl stress up to concentrations of 100 mM, beyond which the plant's growth is negatively impacted. Interestingly, NaCl concentrations exceeding 200 mM do not result in the plant's death but significantly reduce its growth rate (Chen et al., 2023). However, halophytes like *T. ramosissima* have the ability to increase K⁺ retention when under salt stress, facilitating their adaptation to a range of saline and alkaline conditions (Garthwaite et al., 2005). Studies have discovered that the roots of *T. ramosissima* can resist

NaCl stress using 10 mM KCl and can continue to grow even under 200 mM NaCl stress (Chen et al., 2022b; Chen et al., 2022c). In this manuscript, *T. ramosissima* was used as the research subject. We studied and analyzed the role of WRKY genes in *T. ramosissima* in response to NaCl stress when exogenous K⁺ is applied. By delving into the transcriptome, we identified key candidate WRKY genes and their essential metabolic pathways. This provides a scientific theoretical foundation for the selection of salt-tolerant varieties and the involvement of WRKY transcription factors in alleviating the toxic effects of NaCl.

2 Materials and methods

Throughout this study, root samples of *T. ramosissima* under the stress of 200 mM NaCl with the application of exogenous K⁺ at 0, 48, and 168 h were collected for transcriptome sequencing. The data obtained underwent processing, including the screening of Differential expressed genes (DEGs) and database annotation, then identifying the WRKY-related genes. Using Pfam_Scan, protein domains of these genes were predicted (Finn et al., 2016). ProtParam (<https://web.expasy.org/protparam/>) provided predictions on the basic physicochemical properties of the WRKY gene family members (Gasteiger et al., 2003), while CELLOv.2.5 (<http://cello.life.nctu.edu.tw/>) was employed for their subcellular localization. Subsequent analysis focused on the KEGG pathways annotated by the identified WRKY genes. A phylogenetic tree of the key WRKY candidate genes was then constructed. Lastly, the accuracy of the transcriptome data for these genes was validated using Quantitative Real-Time PCR (qRT-PCR) (Supplementary Figure S1).

2.1 Plant materials and treatment

The *T. ramosissima* with similar growth (seedling age: 5 months) were selected. They were cultivated in a 24-well hydroponic box with 1/2 Hoagland nutrient solution (replaced every 3 days), and the experimental treatments were applied 2 months later. The group with 1/2 Hoagland nutrient solution was set as the control group. The seedlings cultivated in 200 mM NaCl 1/2 Hoagland nutrient solution and 200 mM NaCl + 10 mM KCl 1/2 Hoagland nutrient solution were the treatment groups, with 8 plants per group (repeated 3 times). The nutrient solution was replaced every 3 days, with 24 seedlings for all treatments. At the same time, root samples were collected at 0, 48, and 168 h after treatment and stored in a -80°C freezer for transcriptome sequencing.

2.2 Transcriptome sequencing and differentially expressed gene screening

The collected *T. ramosissima* root samples were entrusted to biotechnology company (Gene Denovo, Guangzhou, China) for transcriptome sequencing. During the experiment, ultrasound was used to break down the mRNA. (Note: mRNA is Oligo (dT) magnetic beads enriched to eukaryotic mRNA with a polyA tail).

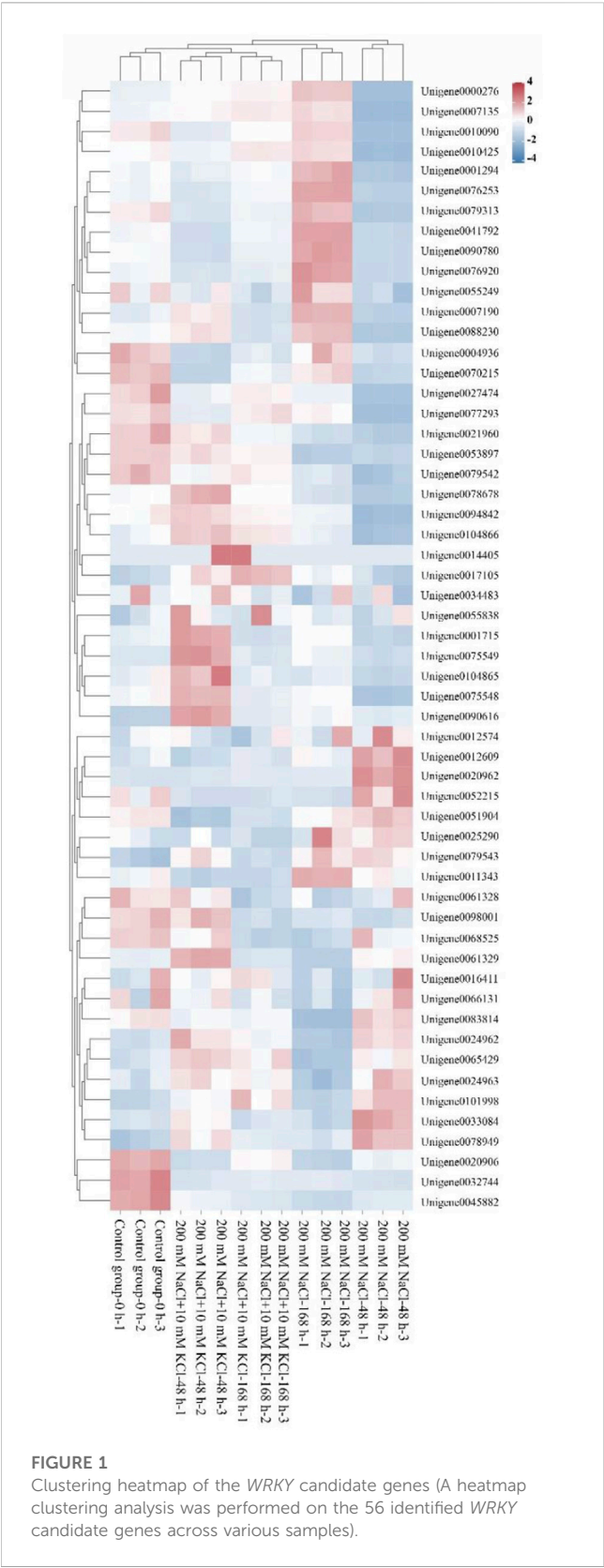
The mRNA (fragmentation) is then used as the basis to synthesize the double strand of cDNA. The purified double-stranded cDNA was screened out using about 200bp and PCR amplification was performed. Finally, the database was constructed using PCR products, which were purified by AMPure XP beads (Chen et al., 2022b). The raw data obtained (WRKY genes in the roots of *T. ramosissima* - NCBI Short Reads Archive database: SRP356215) was analyzed. DEGs were identified based on the criteria of FDR < 0.05 (corrected *p*-value) and $|\log_2FC| > 1$. The DEGs were annotated to the Gene Ontology (GO) (Conesa et al., 2005) and Kyoto Encyclopedia of Genes and Genomes (KEGG) (Michael et al., 2000) databases.

2.3 Prediction of Pfam protein structural domains

Through transcriptome sequencing results, members of the WRKY family annotated in the NCBI database were obtained from the roots of the *T. ramosissima*. Then, the obtained WRKY proteins were matched in the Pfam database using the Pfam_Scan program, yielding protein structure-related annotation information for the WRKY genes (Finn et al., 2016), and the final members of the WRKY gene family were obtained. The basic physicochemical properties of the members of the WRKY gene family were analyzed and predicted online using ProtParam (<https://web.expasy.org/protparam/>), including Molecular weight, Theoretical pI, Grand average of hydropathicity (GRAVY), Instability index, and Aliphatic index, etc. Finally, the obtained WRKY proteins were subjected to subcellular localization prediction analysis using the online website CELLOv.2.5 (<http://cello.life.nctu.edu.tw/>).

2.4 Quantitative Real-Time PCR (qRT-PCR) verification

Seven WRKY candidate genes were randomly chosen to validate the transcriptome sequencing accuracy. Total RNA from *T. ramosissima* root samples was extracted using the RNeasy pure kit (Tiangeng, Beijing, China), and cDNA was synthesized with the Prime Script™ DEG II 1st Strand cDNA Synthesis Kit (Takara, Beijing, China). Primers were designed for the above WRKY genes (Supplementary Table S1). However, after diluting the cDNA template 8 times, qRT-PCR experiments were conducted according to the instructions of the TB Green Premix ExTaq (Tli RNaseH Plus) kit (Takara, RR420A) by Takara company. The reaction system was set to 20 µL, including 10 µL TB Green Premix Ex Taq, 0.8 µL upstream primer (10 µmol L⁻¹), 0.8 µL downstream primer (10 µmol L⁻¹), 1 µL DNA template, and 7.4 µL dd H₂O. Three replicates were set for each sample, with all operations performed on ice. Amplification of each sample was executed using the ABI ViiA™ 7 Real-time PCR system (ABI, California, United States). The PCR amplification protocol included an initial denaturation at 95°C for 30 s, followed by 40 cycles of denaturation at 95°C for 5 s and annealing at 60°C for 30 s. The melting curve process involved heating at 95°C for 5 s, then 60°C for 1 min, gradually increasing to 95°C, and finally cooling to



50°C for 30 s (Yang et al., 2022). Each gene was biologically replicated 3 times, with *Tubulin* as the internal reference gene. The $2^{-\Delta\Delta CT}$ method was used to calculate the relative expression levels (Livak and Schmittgen, 2001).

2.5 Experiment data processing

Data computations such as mean and standard deviations were performed using Microsoft Excel (Microsoft, Washington, United States). The prediction of Pfam protein structural domains was made through the Pfam_Scan program. The construction of graphs was facilitated by the Origin 2019 software (OriginLab, Massachusetts, United States). Phylogenetic trees were developed utilizing the MEGA 11 software (MEGA software, Pennsylvania, United States). Lastly, LSD analysis was executed using the ANOVA tool in the SPSS 26.0 software (SPSS, New York, United States).

3 Results

3.1 Prediction analysis of Pfam protein structure domains

The transcript data results revealed 61 WRKY-related genes. Subsequent predictions of these 61 WRKY gene protein sequences in the Pfam database (Supplementary Table S3) showed that 5 WRKY genes (*Unigene0009588*, *Unigene0016969*, *Unigene0023528*, *Unigene0052366*, and *Unigene0062497*) were not present in the Pfam database. The results show that the start-to-end positions of the structure domains of the Unigene-encoded protein sequences predicted by the HMM models of these 56 WRKY genes are 1-489, the HMM length is all 21-59, the bit score is between 23.7-96.4, the HMM acc is PF03106.15, the clan is CL0274, and the PfamA_definition is WRKY DNA-binding domain. Finally, there are 56 WRKY genes present in the roots of the *T. ramosissima* under the exogenous NaCl stress.

3.2 Identification and physicochemical property analysis of WRKY genes in the roots of *T. ramosissima*

The ProtParam online tool was used to analyze the physicochemical properties of the 56 WRKY genes, and the 56 WRKY genes were analyzed by clustering heat map with each sample (Figure 1). The results show (Table 1) that the number of amino acids ranges from 39 to 728 aa, indicating a large variation in the number of amino acids; the molecular weight ranges from 4385.91 (*Unigene0075549*) to 69860.14 (*Unigene0032744*) Da, which is proportional to the amino acid content; Theoretical pI ranges from 4.70 (*Unigene0077293*) to 10.26 (*Unigene0041792*), among which 27 members are basic proteins (pI > 7), and 29 members are acidic proteins (pI < 7); only 4 WRKY proteins have an Instability index less than 40, indicating that most of them are unstable proteins. The results of hydrophilicity/hydrophobicity of the WRKY proteins obtained are all less than 0, indicating that these 56 WRKY proteins are hydrophilic. The subcellular prediction results show that all WRKY proteins are located in the nuclear.

3.3 Changes in the expression levels of WRKY genes

According to our focus on the expression levels of 56 WRKY genes (Supplementary Figure S2), five genes (*Unigene0066131*,

TABLE 1 WRKY gene family information of *T. ramosissima*

Gene ID	ORF/aa	Molecular weight	Theoretical pI	GRAVY	Instability index	Aliphatic index	Subcellular localization
Unigene0076253	187	21108.72	5.56	-0.921	59.01	27.70	Nuclear
Unigene0061328	51	5689.37	8.62	-0.749	23.26	36.47	Nuclear
Unigene0001294	122	13952.30	9.30	-1.339	56.76	38.36	Nuclear
Unigene0041792	106	11842.45	10.26	-1.024	43.40	38.68	Nuclear
Unigene0070215	114	13023.43	10.19	-1.255	38.85	40.26	Nuclear
Unigene0004936	160	18114.62	5.50	-1.116	38.13	41.31	Nuclear
Unigene0078949	249	28240.02	6.58	-1.168	68.74	45.78	Nuclear
Unigene0012609	232	25655.25	8.76	-0.919	51.71	45.78	Nuclear
Unigene0077293	180	19497.00	4.70	-0.872	46.13	46.72	Nuclear
Unigene0052215	609	67186.50	8.85	-0.863	56.14	48.21	Nuclear
Unigene0021960	355	39213.13	6.40	-0.895	59.03	49.30	Nuclear
Unigene0010090	632	68852.13	6.51	-0.918	53.66	49.56	Nuclear
Unigene0007135	575	62627.58	6.26	-0.774	55.26	49.70	Nuclear
Unigene0020962	252	28622.79	6.77	-0.981	44.02	49.96	Nuclear
Unigene0011343	329	35264.87	6.97	-0.775	62.94	50.46	Nuclear
Unigene0079542	501	53725.20	5.36	-0.854	52.09	50.48	Nuclear
Unigene0007190	255	27861.47	5.92	-0.968	59.44	50.63	Nuclear
Unigene0090616	130	15194.16	9.64	-1.072	41.92	50.92	Nuclear
Unigene0025290	195	22255.82	9.32	-1.010	50.31	51.90	Nuclear
Unigene0098001	566	60735.69	8.46	-0.836	62.84	52.99	Nuclear
Unigene0075548	305	33070.37	5.65	-0.678	57.39	53.18	Nuclear
Unigene0104865	494	53656.32	8.11	-0.845	51.73	53.56	Nuclear
Unigene0032744	665	69860.14	6.23	-0.589	53.48	55.85	Nuclear
Unigene0065429	615	65432.66	5.95	-0.672	49.31	55.97	Nuclear
Unigene0079543	513	56066.51	5.96	-0.743	47.55	56.24	Nuclear
Unigene0079313	550	58821.01	6.22	-0.598	46.49	56.80	Nuclear
Unigene0068525	203	21788.06	9.33	-0.741	48.30	56.80	Nuclear
Unigene0034483	189	21744.23	6.35	-0.953	42.14	57.20	Nuclear
Unigene0075549	39	4385.91	9.24	-0.810	49.69	57.44	Nuclear
Unigene0017105	583	63000.11	6.96	-0.698	49.66	58.04	Nuclear
Unigene0012574	282	32207.34	8.32	-0.883	51.76	58.40	Nuclear
Unigene0104866	506	55331.41	6.93	-0.788	64.38	58.42	Nuclear
Unigene0083814	361	38413.24	9.59	-0.650	45.70	58.67	Nuclear
Unigene0016411	560	61016.73	6.98	-0.832	48.87	60.73	Nuclear
Unigene0033084	375	40205.78	9.17	-0.716	61.11	61.15	Nuclear
Unigene0024963	232	26361.25	8.90	-1.047	33.15	61.25	Nuclear
Unigene0020906	244	27466.33	8.21	-0.981	55.39	61.97	Nuclear
Unigene0061329	370	40649.28	10.02	-0.661	56.21	61.97	Nuclear

(Continued on following page)

TABLE 1 (Continued) WRKY gene family information of *T. ramosissima*

Gene ID	ORF/aa	Molecular weight	Theoretical pI	GRAVY	Instability index	Aliphatic index	Subcellular localization
Unigene0024962	496	54325.27	6.38	−0.900	40.26	62.26	Nuclear
Unigene0014405	140	16207.29	9.92	−0.878	53.83	62.71	Nuclear
Unigene0101998	345	38036.14	5.64	−0.806	49.49	62.75	Nuclear
Unigene0053897	364	40888.56	8.63	−0.830	51.81	62.75	Nuclear
Unigene0045882	581	61880.46	8.22	−0.569	50.31	62.93	Nuclear
Unigene0078678	340	38231.54	6.83	−0.816	45.20	63.94	Nuclear
Unigene0076920	404	43452.25	6.33	−0.475	44.34	64.73	Nuclear
Unigene0055838	280	31828.55	8.21	−0.814	63.52	65.89	Nuclear
Unigene0094842	392	42493.39	5.71	−0.491	48.35	66.71	Nuclear
Unigene0055249	508	54877.26	6.46	−0.623	54.71	67.28	Nuclear
Unigene0010425	152	16724.90	6.73	−0.630	48.43	67.30	Nuclear
Unigene0066131	134	15205.10	9.36	−0.794	60.77	67.61	Nuclear
Unigene0000276	331	36214.41	6.63	−0.663	48.43	67.79	Nuclear
Unigene0027474	390	42967.37	9.71	−0.724	55.91	68.26	Nuclear
Unigene0088230	343	38413.54	6.12	−0.762	47.12	69.39	Nuclear
Unigene0051904	334	36844.91	9.87	−0.645	55.36	71.23	Nuclear
Unigene0090780	175	18904.46	9.40	−0.421	49.35	71.49	Nuclear
Unigene0001715	564	61473.25	7.79	−0.378	47.66	79.95	Nuclear

Note: ORF/aa means Number of amino acids, GRAVY means Grand average of hydropathicity.

Unigene0061329, Unigene0053897, Unigene0025290, and Unigene0101998) showed an initial upregulation followed by downregulation after 48 and 168 h of NaCl stress. However, their expression levels were first downregulated and then upregulated after 48 and 168 h of exogenous K⁺ application under NaCl stress. Another ten genes (Unigene0024962, Unigene00249623, Unigene0079543, Unigene0104866, Unigene0094842, Unigene0078949, Unigene0078678, Unigene0075549, Unigene0065429, and Unigene0033084) showed initial upregulation followed by downregulation under NaCl stress at 48 and 168 h. However, they consistently exhibited an upregulated trend in their expression levels after 48 and 168 h of exogenous K⁺ application under NaCl stress. Notably, Unigene0090616 showed an upregulated trend in expression levels, whether it was under NaCl stress at 48 and 168 h or under NaCl stress with exogenous K⁺ application at 48 and 168 h.

3.4 WRKY genes annotated to KEGG pathway

Ten DEGs of the 56 WRKY genes found in the roots of *T. ramosissima* were annotated in the related KEGG pathway (Supplementary Table S2) when exogenous K⁺ was applied at 0, 48 and 168 h under NaCl stress. 6 WRKY genes (Unigene0007135, Unigene0010090, Unigene0014405, Unigene0052215, Unigene0070215 and Unigene0077293) were annotated in the plant-interaction pathway (ko04626) and the MAPK signaling pathway-plant pathway (ko04016). The 6 WRKY genes were

classified into Organismal Systems, and the Environmental Information Processing in KEGG A. KEGG B classified Environmental adaptation and Signal transduction. Unigene0024962, Unigene0024963, Unigene0079542 and Unigene0079543 are annotated for the Plant-pathogen interaction pathway (ko04626). In addition, these 4 WRKY genes were classified into Organismal Systems in KEGG A and Environmental adaptation in KEGG B. According to the expression levels of the 10 WRKY genes at 0, 48 and 168 h under NaCl stress (Supplementary Figure S2), The expression levels of 3 WRKY genes (Unigene0024962, Unigene0024963, and Unigene0079543) increased at 48 h and 168 h when applying exogenous K⁺ under NaCl stress. Notably, the expression levels of Unigene0024962 and Unigene0024963 consistently rose at both 48 and 168 h with exogenous K⁺ under NaCl stress. However, under NaCl stress alone, their expression levels first increased at 48 h and then decreased at 168 h.

3.5 Analysis of plant-pathogen interaction pathway

As inferred from the analysis of the Plant-pathogen interaction pathway (Table 2), In the 200 mM NaCl-48 h vs. 200 mM NaCl + 10 mM KCl-48 h (N-48 h vs. N + K-48 h) comparison group, 99 DEGs were significantly enriched into the Plant-pathogen interaction pathway ($p < 0.05$). Among them, 65 DEGs were upregulated, while 34 were downregulated. The 7 WRKY genes (Unigene0052215, Unigene0010090, Unigene0007135, Unigene0070215, Unigene0077293, Unigene0024962

TABLE 2 Analysis of Plant-pathogen interaction pathway.

Pathway	Gene numbers	Class	p-Value	Up	Down
N-48 h vs. N + K-48 h					
Plant-pathogen interaction	99	Organismal Systems	0.013055	65	34
N-168 h vs. N + K-168 h					
Plant-pathogen interaction	74	Organismal Systems	0.186311	21	53

Note: N-48 h: 200 mM NaCl-48 h, N + K-48 h: 200 mM NaCl + 10 mM KCl-48 h, N-168 h: 200 mM NaCl-168 h, N + K-168 h: 200 mM NaCl + 10 mM KCl-168 h.

TABLE 3 Analysis of MAPK signaling pathway-plant pathway.

Pathway	Gene numbers	Class	p-Value	Up	Down
N-48 h vs. N + K-48 h					
MAPK signaling pathway-plant pathway	57	Environmental Information Processing	0.297932	37	20
N-168 h vs. N + K-168 h					
MAPK signaling pathway-plant pathway	54	Environmental Information Processing	0.067500	19	35

Note: N-48 h: 200 mM NaCl-48 h, N + K-48 h: 200 mM NaCl + 10 mM KCl-48 h, N-168 h: 200 mM NaCl-168 h, N + K-168 h: 200 mM NaCl + 10 mM KCl-168 h.

and *Unigene00249623*) have been annotated for Plant-pathogen interaction pathway in N-48 h vs. N + K-48 h comparison group, and the expression level changed (Supplementary Figure S3). In addition, *Unigene0052215*, *Unigene0024962*, and *Unigene00249623* were down-regulated, while *Unigene0010090*, *Unigene0007135*, *Unigene0070215*, and *Unigene0077293* were upregulated, suggesting their roles in NaCl stress resistance. In the comparison of the 200 mM NaCl-168 h and 200 mM NaCl + 10 mM KCl-168 h groups, 74 DEGs were identified in the Plant-pathogen interaction pathway. Of these, 21 DEGs were upregulated and 53 were downregulated. Notably, 3 WRKY genes (*Unigene0070215*, *Unigene0024962*, and *Unigene00249623*) were associated with this pathway. The expression level of *Unigene0070215* was downregulated, while the expression levels of *Unigene0024962* and *Unigene00249623* were upregulated.

3.6 Analysis of MAPK signaling pathway-plant pathway

Based on the MAPK signaling pathway-plant pathway (Table 3), in the 200 mM NaCl-48 h vs. 200 mM NaCl + 10 mM KCl-48 h comparison group, 57 DEGs were enriched to MAPK signaling Pathway-plant pathway. Among them, 37 DEGs showed upregulated expression levels, while 20 DEGs exhibited downregulated expression. *Unigene0052215*, *Unigene0010090*, *Unigene0007135*, *Unigene0070215* and *Unigene0077293* belong to WRKY genes that have been annotated to MAPK signaling the N-48 h vs. N + K-48 h comparison group Pathway-plant pathway, and the expression level was changed (Supplementary Figure S4). Among them, in addition to the downregulated expression of *Unigene0052215*, the upregulated expression of *Unigene0010090*, *Unigene0007135*, *Unigene0070215* and *Unigene0077293* actively participate in the resistance to NaCl stress. In the N-168 h vs. N + K-168 h comparison group, 54 DEGs were significantly enriched to MAPK signaling Pathway-plant pathway ($p < 0.05$). Among

them, 19 DEGs showed upregulated expression levels, while 35 DEGs exhibited downregulated expression. Only *Unigene0070215* belongs to the WRKY gene and has been annotated to MAPK signaling pathway-plant in N-168 h vs. N + K-168 h comparison group pathway, and its expression level was still downregulated.

3.7 Candidate key WRKY genes were constructed for phylogenetic tree analysis

Unigene0024962 and *Unigene00249623*, which belong to the WRKY gene family (WRKY transcription factor 1), have both been annotated in the Plant-pathogen interaction pathway. Under NaCl stress, their expression levels initially increase at 48 h and then decrease at 168 h. However, after exogenous K⁺ application under NaCl stress, their expression levels continue to show an increasing trend at both 48 and 168 h. Thus, it can be inferred that *Unigene0024962* and *Unigene00249623* are key candidate genes in the WRKY family. The protein amino acid sequences of *Unigene0024962* and *Unigene0024963* were compared on the NCBI using Blast, resulting in 15 homologous species were selected (Supplementary Table S4). A phylogenetic tree constructed with MEGA software revealed that *Unigene0024962* shares a close phylogenetic relationship with *Tamarix hispida* (Supplementary Figure S5).

3.8 Quantitative Real-Time PCR (qRT-PCR) validation of candidate WRKY genes

Seven WRKY candidate genes were selected for qRT-PCR validation to verify the accuracy of our transcriptome sequencing outcomes. The patterns observed in qRT-PCR expression closely matched those from the transcriptomic study (Supplementary Figure S6). This affirms the integrity and dependability of the

transcriptome data acquired. Hence, this research lays a solid foundation for identifying WRKY genes in *T. ramosissima* roots that enhance salinity resilience and mitigate the adverse impacts of NaCl stress.

4 Discussion

Tamarix plants has developed a complex regulatory network to adapt to abiotic stress (Zhu, 2016). Transcription factors are the most important regulatory factors in all abiotic stress responses (Yao et al., 2016). The WRKY family is large, *Arabidopsis thaliana* had 74 WRKY genes (Eulgem et al., 2000), cotton had 116 WRKY genes (Dou et al., 2014), rice (*Oryza sativa*) had 102 WRKY genes (Ross et al., 2007), *Solanum Lycopersicum* had 81 WRKY genes (Huang et al., 2012), carrot (*Daucus carota* L.) had 95 WRKY genes (Li et al., 2016), and Castor Bean (*Ricinus communis* L.) had 58 WRKY genes (Zou et al., 2016). In this study, 56 WRKY genes were found in the roots of *T. ramosissima* under NaCl stress at 48 h and 168 h. Further analysis of these 56 WRKY genes can help to understand the response mechanism of *T. ramosissima* to NaCl stress under exogenous K⁺ application. To identify salt tolerance genes and related metabolic pathways to provide a scientific and theoretical basis for further research on salt tolerance in *T. ramosissima*.

WRKY transcription factors are essential in plant responses to biotic and abiotic stresses (Jiang et al., 2017). Particularly, it plays a crucial role in responding to salt stress in various plants (Huang et al., 2022). A WRKY transcription factor *PbWRKY40* from *Pyrus betulaefolia* functions positively in salt tolerance and modulating organic acid accumulation by regulating *PbVHA-B1* expression (Lin et al., 2022). *Ahwrky75* was overexpressed in the M34 mutant of “salt-tolerant” peanut, which improved the salt tolerance of transgenic peanut (Zhu et al., 2021). In transgenic tobacco, overexpression of the WRKY transcription factor gene of *Begonia* (*MbWRKY5*) can improve drought tolerance and salt tolerance (Han et al., 2018). In *Solanum lycopersicum*, *SlWRKY8* plants overexpressed showed tolerance to salt stress (Gao et al., 2020). In this study, 37 DEGs were upregulated, and 24 DEGs were downregulated in the N-48 h vs. N + K-48 h comparison group. In the N-168 h vs. N + K-168 h comparison group, 33 DEGs were upregulated, and 28 DEGs were downregulated. Additionally, overexpression of the WRKY gene *DgWRKY1* or *DgWRKY3* in chrysanthemum plants increased their salt tolerance by *dendranthema grandiflorum* (Liu et al., 2013). However, K⁺ can better regulate Na⁺ uptake under salt conditions, which plays a vital role in regulating plant physiological processes to enable plants to survive under stress conditions, especially in improving their tolerance to salt stress (Çolpan et al., 2013). Furthermore, the introduction of K⁺ helped mitigate the negative impacts of Na⁺, enhanced the uptake of K⁺, and raised the K⁺/Na⁺ ratio under conditions of NaCl stress (Carden et al., 2003). It is worth noting that halophytes have a solid competitive advantage in maintaining K⁺ stability under a high NaCl stress (Ardie et al., 2009). In this study, expression level at 48 and 168 h increased first and then decreased under NaCl stress. However, the expression level of the candidate key WRKY genes (*Unigene0024962* and *Unigene0024963*) were upregulated after the application of exogenous K⁺ for 48 and

168 h under NaCl stress. Therefore, the increased expression level of *Unigene0024962* and *Unigene0024963* may be influenced by the addition of exogenous K⁺. In brief, the results showed that WRKY genes expression in the roots of *T. ramosissima* was consistently upregulated within 168 h after applying exogenous NaCl stress, which improved the tolerance and resistance of *T. ramosissima* to NaCl stress.

Under conditions of salt stress in grapes (*Vitis vinifera*), *VvWRKY28* has the ability to boost the accumulation levels of enzymatic antioxidants (superoxide dismutase, peroxidase, and catalase) as well as proline. These changes counteract the over-accumulation of reactive oxygen species and support osmotic balance in cells, thus increasing the plant's overall salt tolerance (Liu et al., 2022). In *T. ramosissima*, when exposed to 200 mM NaCl stress for 48 h and 168h, the leaves had more activity of certain helpful enzymes. At the same time, there was more proline in the roots. These changes helped the seeds handle salt better. (Chen et al., 2022b; Chen et al., 2022c). In this study, the expression level of the WRKY gene (*Unigene0020906*) was upregulated after applying exogenous K⁺ for 48 and 168 h under 200 mM NaCl stress. Therefore, it may play an essential role in increasing the main antioxidant activities and proline content accumulation of *T. ramosissima*. Besides, abiotic stresses such as high temperature, salinity and drought affect plant growth. They also affect plant defense response to pathogens through enhancement or inhibition of pathogens, to regulate plant response to pathogen infection (Pandey and Senthil-Kumar, 2019). In particular, one of the earliest signaling events after a plant sense an invading pathogen is the activation of mitogen activated protein kinase (MAPK) signaling cascades. They are able to correlate the perception of external stimuli with cellular responses, and MAPK signaling networks play specific and overlapping roles in controlling the activity of a large number of transcription factors, enzymes, hormones, peptides, and antibacterial chemicals, among others (Taj et al., 2014). WRKY transcription factors have been shown to regulate plant defense responses (Rushton and Somssich, 1998). WRKY transcription factors regulate plant defense responses (Rushton and Somssich, 1998). Their activity is modulated by phosphorylation through mitogen-activated protein kinase (Ishihama and Yoshioka, 2012; Adachi et al., 2015). WRKY protein is involved in *Arabidopsis thaliana* response to bacterial flagellin flg22 as a transcriptional activator at the end of the PAMP signaling cascade. Flagellin flg22 in *Arabidopsis thaliana* initiates the entire MEKK1-MKK4/5-MAPK3/6 MAPK cascade pathway, leading to the activation of WRKY22/29 expressions. These expressions function after the flagellin receptor FLS2, a leucine-rich repeat (LRR) receptor kinase, and further stimulate MAPK cascades, enhancing plant resistance (Sheen et al., 2002). Simultaneously, AtMKK1 and AtMKK2 in *Arabidopsis Thaliana* are also associated with biological and abiotic stress responses as part of a signaling cascade that includes MEKK1 and MPK4 (Teige et al., 2004). Nevertheless, AtMKK2 is also involved in mediating salt signaling (Hadiarto et al., 2006). It is worth noting that Mkk1 is a positive regulator of flg22 activation of MPK3, MPK4, MPK6, and a negative regulator of flg22 activation of gene expression, and its role is very complex (Meszaros et al., 2006). MPK3 and MPK6 are two mitogen-activated protein kinases (MAPK or

MPK) that play key roles in plant resistance by regulating various defense responses (Meng et al., 2013). *AtWRKY33* mutants in *Arabidopsis Thaliana* are very sensitive to NaCl, and overexpression of *AtWRKY33* can enhance tolerance to NaCl stress (Jiang and Deyholos, 2009; Li et al., 2011). In this study, FLS2 activated downstream mitogen-activated protein kinase (MEKK) in both the Plant-pathogen interaction pathway and the MAPK signaling Pathway-plant pathway, MKK1/MKK2 → MPK4 or MKK4/MKK5 → MPK3/MPK6 (+p is the process of protein phosphorylation). They transmit these signals into the nucleus to transcription factors (WRKY 22/WRKY29 or WRKY25/WRKY33) that activate the transcription of genes associated with immune defense. The expression levels of *Unigene0010090*, *Unigene0007135*, *Unigene0070215* and *Unigene0077293* were upregulated in the N-48 h vs. N + K-48 h comparison group. In the comparison group of N-168 h vs. N + K-168 h, the expression levels of *Unigene0024962* and *Unigene0024963* were upregulated. Therefore, it is concluded that these genes actively participate in the resistance to NaCl stress when exogenous K⁺ is applied for 48 h under NaCl stress. However, with the increase of K⁺ addition time, the damage of NaCl stress was alleviated, *T. ramosissima* could maintain normal growth, and the genes involved in regulation were reduced.

To conclude, the results of this manuscript suggest that the result of this manuscript suggest that the WRKY transcription factors are involved in *T. ramosissima*'s response to external K⁺ under NaCl stress. A significant number of WRKY-related genes are upregulated and take part in essential metabolic pathways, thereby creating a comprehensive defense mechanism to help *T. ramosissima* resist NaCl stress. Notably, two WRKY candidate genes, *Unigene0024962* and *Unigene0024963*, play pivotal roles within these pathways. They are identified as key WRKY genes worthy of further investigation.

5 Conclusion

The WRKY transcription factor family is one of the largest transcriptional regulatory families in higher plants, involved in regulating plant growth and development, and responding to biotic and abiotic stresses. They play an indispensable role in the normal life activities of plants, being involved in various aspects of the plant's lifecycle. In this study, the roots of *T. ramosissima*, under exogenous potassium in response to salt stress at 48 h and 168 h, saw the involvement of the WRKY transcription factor family in the growth and development processes of *T. ramosissima*, as well as the construction of its defense system. A significant number of WRKY genes actively upregulated their expression levels to resist NaCl stress through the plant-pathogen interaction pathway and the MAPK signaling pathway-plant pathway, alleviating the damage caused by NaCl and thus maintaining the normal growth of *T. ramosissima*. Notably, two key candidate WRKY genes (*Unigene0024962* and *Unigene0024963*) actively upregulated their expression levels in the plant-pathogen interaction pathway after exogenous potassium application under NaCl stress, playing a crucial role in the *T. ramosissima* root system's mitigation of NaCl stress. They can serve as key salt tolerance genes in the *Tamarix* plants for further exploration.

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/>, SRP356215.

Author contributions

YC: Data curation, Investigation, Project administration, Resources, Software, Writing-review and editing, Methodology, Validation, Conceptualization, Formal Analysis, Visualization, Writing-original draft. XZ: Data curation, Investigation, Methodology, Writing-original draft. YF: Software, Writing-review and editing, Data curation. DS: Project administration, Writing-review and editing, Funding acquisition. JJ: Data curation, Software, Writing-review and editing, Investigation, Resources, Supervision. LW: Investigation, Project administration, Resources, Validation, Writing-review and editing, Funding acquisition, Supervision.

Funding

The authors declare financial support was received for the research, authorship, and/or publication of this article. This work was jointly supported by grants from Jiangsu Coastal Tidal *Spartina alterniflora* Loisel Comprehensive Management Technology Innovation and Demonstration (Su[2023]TG10).

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 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.

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/fgene.2023.1274288/full#supplementary-material>

References

- Adachi, H., Nakano, T., Miyagawa, N., Ishihama, N., Yoshioka, M., Katou, Y., et al. (2015). WRKY transcription factors phosphorylated by MAPK regulate a plant immune NADPH oxidase in *Nicotiana benthamiana*. *Plant Cell* 27, 2645–2663. doi:10.1105/tpc.15.00213
- Amtmann, A., Trouillard, S., and Armengaud, P. (2008). The effect of potassium nutrition on pest and disease resistance in plants. *Physiol. Plant* 133, 682–691. doi:10.1111/j.1399-3054.2008.01075.x
- Ardie, S. W., Xie, L., Takahashi, R., Liu, S., and Takano, T. (2009). Cloning of a high-affinity K⁺ transporter gene *PutHKT2;1* from *Puccinellia tenuiflora* and its functional comparison with *OsHKT2;1* from rice in yeast and *Arabidopsis*. *J. Exp. Bot.* 60, 3491–3502. doi:10.1093/jxb/erp184
- Bahramsoltani, R., Kalkhorani, M., Abbas, Z. S., Farzaei, M. H., and Rahimi, R. (2020). The genus *Tamarix*: traditional uses, phytochemistry, and pharmacology. *J. Ethnopharmacol.* 246, 112245. doi:10.1016/j.jep.2019.112245
- Boyer, J. S. (1982). Plant productivity and environment. *Science* 218, 443–448. doi:10.1126/science.218.4571.443
- Carden, D. E., Walker, D. J., Flowers, T. J., and Miller, A. J. (2003). Single-cell measurements of the contributions of cytosolic Na⁺ and K⁺ to salt tolerance. *Plant Physiol.* 131, 676–683. doi:10.1104/pp.011445
- Chakraborty, K., Bhaduri, D., Meena, H. N., and Kalariya, K. (2016). External potassium K⁺ application improves salinity tolerance by promoting Na⁺-exclusion, K⁺-accumulation and osmotic adjustment in contrasting peanut cultivars. *Plant Physiol. Biochem.* 103, 143–153. doi:10.1016/j.plaphy.2016.02.039
- Che, B., Cheng, C., Fang, J., Liu, Y., Jiang, L., and Yu, B. (2019). The recretehalophyte *Tamarix TrSOS1* gene confers enhanced salt tolerance to transgenic hairy root composite cotton seedlings exhibiting Virus-Induced gene silencing of *GhSOS1*. *Int. J. Mol. Sci.* 20, 2930. doi:10.3390/ijms20122930
- Chen, Y., Li, H., Zhang, S., Du, S., Zhang, J., Song, Z., et al. (2023). Analysis of the main antioxidant enzymes in the roots of *Tamarix ramosissima* under NaCl stress by applying exogenous potassium (K⁺). *Front. Plant Sci.* 14, 1114266. doi:10.3389/fpls.2023.1114266
- Chen, Y., Zhang, S., Du, S., Jiang, J., and Wang, G. (2022a). Transcriptome and metabolomic analysis of *Tamarix ramosissima* potassium (K⁺) Channels and transporters in response to NaCl Stress. *Genes* 13, 1313. doi:10.3390/genes13081313
- Chen, Y., Zhang, S., Du, S., Zhang, X., Jiang, J., and Wang, G. (2022b). Analysis of amino acids in the roots of *Tamarix ramosissima* by application of exogenous potassium (K⁺) under NaCl Stress. *Int. J. Mol. Sci.* 23, 9331. doi:10.3390/ijms23169331
- Chen, Y., Zhang, S., Du, S., Zhang, X., Wang, G., Huang, J., et al. (2022c). Effects of exogenous potassium (K⁺) application on the antioxidant enzymes activities in leaves of *Tamarix ramosissima* under NaCl Stress. *Genes* 13, 1507. doi:10.3390/genes13091507
- Cheng, Z., Luan, Y., Meng, J., Sun, J., Tao, J., and Zhao, D. (2021). WRKY transcription factor response to High-Temperature stress. *Plants (Basel)* 10, 2211. doi:10.3390/plants10102211
- Çolpan, E., Zengin, M., and Özbahçe, A. (2013). The effects of potassium on the yield and fruit quality components of stick tomato. *Hortic. Environ. Biotechnol.* 54, 20–28. doi:10.1007/s13580-013-0080-4
- Conesa, A., Gotz, S., Garcia-Gomez, J. M., Terol, J., Talon, M., and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21, 3674–3676. doi:10.1093/bioinformatics/bti610
- Dai, W., Wang, M., Gong, X., and Liu, J. H. (2018). The transcription factor FcWRKY40 of *Fortunella crassifolia* functions positively in salt tolerance through modulation of ion homeostasis and proline biosynthesis by directly regulating SOS2 and P5CS1 homologs. *New Phytol.* 219, 972–989. doi:10.1111/nph.15240
- Dou, L., Zhang, X., Pang, C., Song, M., Wei, H., Fan, S., et al. (2014). Genome-wide analysis of the WRKY gene family in cotton. *Mol. Genet. Genomics* 289, 1103–1121. doi:10.1007/s00438-014-0872-y
- Duan, Q., Zhu, Z., Wang, B., and Chen, M. (2022). Recent progress on the salt tolerance mechanisms and application of tamarisk. *Int. J. Mol. Sci.* 23, 3325. doi:10.3390/ijms23063325
- Eulgem, T., Rushton, P. J., Robatzek, S., and Somssich, I. E. (2000). The WRKY superfamily of plant transcription factors. *Trends Plant Sci.* 5, 199–206. doi:10.1016/s1360-1385(00)01600-9
- Finn, R. D., Coghill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., et al. (2016). The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* 44, D279–D285. doi:10.1093/nar/gkv1344
- Flowers, T. J., and Colmer, T. D. (2008). Salinity tolerance in halophytes. *New Phytologist* 179, 945–963. doi:10.1111/j.1469-8137.2008.02531.x
- Gao, C., Wang, Y., Jiang, B., Liu, G., Yu, L., Wei, Z., et al. (2011). A novel vacuolar membrane H⁺-ATPase c subunit gene (*ThVHAcl*) from *Tamarix hispida* confers tolerance to several abiotic stresses in *Saccharomyces cerevisiae*. *Mol. Biol. Rep.* 38, 957–963. doi:10.1007/s11033-010-0189-9
- Gao, Y. F., Liu, J. K., Yang, F. M., Zhang, G. Y., Wang, D., Zhang, L., et al. (2020). The WRKY transcription factor WRKY8 promotes resistance to pathogen infection and mediates drought and salt stress tolerance in *Solanum lycopersicum*. *Physiol. Plant* 168, 98–117. doi:10.1111/ppl.12978
- Garthwaite, A. J., von Bothmer, R., and Colmer, T. D. (2005). Salt tolerance in wild *Hordeum* species is associated with restricted entry of Na⁺ and Cl⁻ into the shoots. *J. Exp. Bot.* 56, 2365–2378. doi:10.1093/jxb/eri229
- Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D., and Bairoch, A. (2003). ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* 31, 3784–3788. doi:10.1093/nar/gkg563
- Gupta, B., Huang, B., and Catuvelli, L. (2014). Mechanism of salinity tolerance in plants: physiological, biochemical, and molecular characterization. *Int. J. Genomics* 2014, 1–18. doi:10.1155/2014/701596
- Hadiarto, T., Nanmori, T., Matsuoka, D., Iwasaki, T., Sato, K., Fukami, Y., et al. (2006). Activation of *Arabidopsis* MAPK kinase kinase (AtMEKK1) and induction of AtMEKK1-AtMEK1 pathway by wounding. *Planta* 223, 708–713. doi:10.1007/s00425-005-0126-7
- Han, D., Hou, Y., Wang, Y., Ni, B., Li, Z., and Yang, G. (2018). Overexpression of a *Malus baccata* WRKY transcription factor gene (*MbWRKY5*) increases drought and salt tolerance in transgenic tobacco. *Can. J. Plant Sci.* 99, 173–183. doi:10.1139/cjps-2018-0053
- Hao, S., Wang, Y., Yan, Y., Liu, Y., Wang, J., and Chen, S. (2021). A review on plant responses to salt stress and their mechanisms of salt resistance. *Horticulturae* 7, 132. doi:10.3390/horticulturae7060132
- Huang, J., Liu, F., Chao, D., Xin, B., Liu, K., Cao, S., et al. (2022). The WRKY transcription factor OsWRKY54 is involved in salt tolerance in rice. *Int. J. Mol. Sci.* 23, 11999. doi:10.3390/ijms231911999
- Huang, S., Gao, Y., Liu, J., Peng, X., Niu, X., Fei, Z., et al. (2012). Genome-wide analysis of WRKY transcription factors in *Solanum lycopersicum*. *Mol. Genet. Genomics* 287, 495–513. doi:10.1007/s00438-012-0696-6
- Ishihama, N., and Yoshioka, H. (2012). Post-translational regulation of WRKY transcription factors in plant immunity. *Curr. Opin. Plant Biol.* 15, 431–437. doi:10.1016/j.pbi.2012.02.003
- Jiang, J., Ma, S., Ye, N., Jiang, M., Cao, J., and Zhang, J. (2017). WRKY transcription factors in plant responses to stresses. *J. Integr. Plant Biol.* 59, 86–101. doi:10.1111/jipb.12513
- Jiang, Y., and Deyholos, M. K. (2009). Functional characterization of *Arabidopsis* NaCl-inducible WRKY25 and WRKY33 transcription factors in abiotic stresses. *Plant Mol. Biol.* 69, 91–105. doi:10.1007/s11103-008-9408-3
- Johnson, R., Vishwakarma, K., Hossen, M. S., Kumar, V., Shackira, A. M., Puthur, J. T., et al. (2022). Potassium in plants: growth regulation, signaling, and environmental stress tolerance. *Plant Physiol. Biochem.* 172, 56–69. doi:10.1016/j.plaphy.2022.01.001
- Khan, S. A., Li, M. Z., Wang, S. M., and Yin, H. J. (2018). Revisiting the role of plant transcription factors in the battle against abiotic stress. *Int. J. Mol. Sci.* 19, 1634. doi:10.3390/ijms19061634
- Leigh, R. A., and Jones, R. G. W. (1984). A hypothesis relating critical potassium concentrations for growth to the distribution and functions of this ion in the plant cell. *New Phytol.* 97, 1–13. doi:10.1111/j.1469-8137.1984.tb04103.x
- Li, J., Pu, L., Han, M., Zhu, M., Zhang, R., and Xiang, Y. (2014). Soil salinization research in China: advances and prospects. *J. Geogr. Sci.* 24, 943–960. doi:10.1007/s11442-014-1130-2
- Li, K., and Zhou, C. J. (2014). Research progress in WRKY transcription factors in plants. *Plant Physiology J.* 50, 1329–1335. doi:10.13592/j.cnki.pj.2014.1019
- Li, M. Y., Xu, Z. S., Tian, C., Huang, Y., Wang, F., and Xiong, A. S. (2016). Genomic identification of WRKY transcription factors in carrot (*Daucus carota*) and analysis of evolution and homologous groups for plants. *Sci. Rep.* 6, 23101–23117. doi:10.1038/srep23101
- Li, S., Fu, Q., Chen, L., Huang, W., and Yu, D. (2011). *Arabidopsis thaliana* WRKY25, WRKY26, and WRKY33 coordinate induction of plant thermotolerance. *Planta* 233, 1237–1252. doi:10.1007/s00425-011-1375-2
- Lin, L., Yuan, K., Huang, Y., Dong, H., Qiao, Q., Xing, C., et al. (2022). A WRKY transcription factor *PbWRKY40* from *Pyrus betulaefolia* functions positively in salt tolerance and modulating organic acid accumulation by regulating *PbVHA-B1* expression. *Environ. Exp. Bot.* 196, 104782. doi:10.1016/j.envexpbot.2022.104782
- Liu, J. N., Fang, H., Liang, Q., Dong, Y., Wang, C., Yan, L., et al. (2022a). Genomic analyses provide insights into the evolution and salinity adaptation of halophyte *Tamarix chinensis*. *Gigascience* 12, giad053. doi:10.1093/gigascience/giad053
- Liu, Q. L., Zhong, M., Li, S., Pan, Y. Z., Jiang, B. B., Jia, Y., et al. (2013). Overexpression of a chrysanthemum transcription factor gene, *DgWRKY3*, in tobacco enhances tolerance to salt stress. *Plant Physiol. Biochem.* 69, 27–33. doi:10.1016/j.plaphy.2013.04.016
- Liu, W., Liang, X., Cai, W., Wang, H., Liu, X., Cheng, L., et al. (2022b). Isolation and functional analysis of *VvWRKY28*, a *vitis vinifera* WRKY transcription factor gene, with functions in tolerance to cold and salt stress in transgenic *Arabidopsis thaliana*. *Int. J. Mol. Sci.* 23, 13418. doi:10.3390/ijms232113418
- Liu, X., Song, Y., Xing, F., Wang, N., Wen, F., and Zhu, C. (2016). *GhWRKY25*, a group I WRKY gene from cotton, confers differential tolerance to abiotic and biotic

- stresses in transgenic *Nicotiana benthamiana*. *Protoplasma* 253, 1265–1281. doi:10.1007/s00709-015-0885-3
- Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using Real-Time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods* 25, 402–408. doi:10.1006/meth.2001.1262
- Ludwig, M., Wilmes, P., and Schrader, S. (2018). Measuring soil sustainability via soil resilience. *Sci. Total Environ.* 626, 1484–1493. doi:10.1016/j.scitotenv.2017.10.043
- Mauro, R. P., Perez-Alfocea, F., Cookson, S. J., Ollat, N., and Vitale, A. (2022). Editorial: physiological and molecular aspects of plant Rootstock-Scion interactions. *Front. Plant Sci.* 13, 852518. doi:10.3389/fpls.2022.852518
- Meng, X., Xu, J., He, Y., Yang, K., Mordorski, B., Liu, Y., et al. (2013). Phosphorylation of an ERF transcription factor by *Arabidopsis* MPK3/MPK6 regulates plant defense gene induction and fungal resistance. *Plant Cell* 25, 1126–1142. doi:10.1105/tpc.112.109074
- Meszáros, T., Helfer, A., Hatzimasoura, E., Magyar, Z., Serazetdinova, L., Rios, G., et al. (2006). The *Arabidopsis* MAP kinase kinase MKK1 participates in defence responses to the bacterial elicitor flagellin. *Plant J.* 48, 485–498. doi:10.1111/j.1365-3113.2006.02888.x
- Michael, A., Catherine, A. B., Judith, A. B., David, B., Heather, B., Cherry, J. M., et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* 25, 25–29. doi:10.1038/75556
- Munns, R., and Tester, M. (2008). Mechanisms of salinity tolerance. *Annu. Rev. Plant Biol.* 59, 651–681. doi:10.1146/annurev.arplant.59.032607.092911
- Pandey, P., and Senthil-Kumar, M. (2019). Plant-pathogen interaction in the presence of abiotic stress: what do we know about plant responses? *Indian J. Plant Physiology* 24, 541–549. doi:10.1007/s40502-019-00483-7
- Park, H. J., Kim, W. Y., and Yun, D. J. (2013). A role for GIGANTEA: keeping the balance between flowering and salinity stress tolerance. *Plant Signal Behav.* 8, e24820. doi:10.4161/psb.24820
- Qadir, M., Quillérrou, E., Nangia, V., Murtaza, G., Singh, M., Thomas, R., et al. (2014). Economics of salt-induced land degradation and restoration. *Nat. Resour. Forum* 38, 282–295. doi:10.1111/1477-8947.12054
- Ross, C. A., Liu, Y., and Shen, Q. J. (2007). The WRKY gene family in rice (*Oryza sativa*). *J. Integr. Plant Biol.* 49, 827–842. doi:10.1111/j.1744-7909.2007.00504.x
- Rushton, P. J., and Somssich, I. E. (1998). Transcriptional control of plant genes responsive to pathogens. *Curr. Opin. Plant Biol.* 1, 311–315. doi:10.1016/1369-5266(88)80052-9
- Savary, S., Akter, S., Almekinders, C., Harris, J., Korsten, L., Rötter, R., et al. (2020). Mapping disruption and resilience mechanisms in food systems. *Food Secur.* 12, 695–717. doi:10.1007/s12571-020-01093-0
- Schluttenhofer, C., and Yuan, L. (2015). Regulation of specialized metabolism by WRKY transcription factors. *Plant Physiol. (Bethesda)* 167, 295–306. doi:10.1104/pp.114.251769
- Shabala, S., and Cuin, T. A. (2008). Potassium transport and plant salt tolerance. *Physiol. Plant.* 133, 651–669. doi:10.1111/j.1399-3054.2007.01008.x
- Sheen, J., Asai, T., Tena, G., Plotnikova, J., Willmann, M. R., Gomez-Gomez, L., et al. (2002). MAP kinase signalling cascade in *Arabidopsis* innate immunity. *Nat. Lond.* 415, 977–983. doi:10.1038/415977a
- Singh, K., Foley, R. C., and Onate-Sanchez, L. (2002). Transcription factors in plant defense and stress responses. *Curr. Opin. Plant Biol.* 5, 430–436. doi:10.1016/s1369-5266(02)00289-3
- Song, Y., Li, J., Sui, Y., Han, G., Zhang, Y., Guo, S., et al. (2020). The sweet sorghum *SbWRKY50* is negatively involved in salt response by regulating ion homeostasis. *Plant Mol. Biol.* 102, 603–614. doi:10.1007/s11103-020-00966-4
- Sun, J., Zhao, X., Fang, Y., Xu, W., Gao, F., Zhao, W., et al. (2022). Root growth and architecture of *Tamarix chinensis* in response to the groundwater level in the Yellow River Delta. *Mar. Pollut. Bull.* 179, 113717. doi:10.1016/j.marpolbul.2022.113717
- Taj, G., Giri, P., Kumar, A., and Tasleem, M. (2014). MAPK signaling cascades and transcriptional reprogramming in plant-pathogen interactions. *Approaches Plant Stress their Manag.* 297–316. doi:10.1007/978-81-322-1620-9_17
- Tao, Z., Kou, Y., Liu, H., Li, X., Xiao, J., and Wang, S. (2011). OsWRKY45 alleles play different roles in abscisic acid signalling and salt stress tolerance but similar roles in drought and cold tolerance in rice. *J. Exp. Bot.* 62, 4863–4874. doi:10.1093/jxb/err144
- Teige, M., Scheikl, E., Eulgem, T., Dóczi, R., Ichimura, K., Shinozaki, K., et al. (2004). The MKK2 pathway mediates cold and salt stress signaling in *Arabidopsis*. *Mol. Cell* 15, 141–152. doi:10.1016/j.molcel.2004.06.023
- Tittal, M., Mir, R. A., Jatav, K. S., and Agarwal, R. M. (2021). Supplementation of potassium alleviates water stress-induced changes in *Sorghum bicolor* L. *Physiol. Plant* 172, 1149–1161. doi:10.1111/pp.13306
- Ulker, B., and Somssich, I. E. (2004). WRKY transcription factors: from DNA binding towards biological function. *Curr. Opin. Plant Biol.* 7, 491–498. doi:10.1016/j.pbi.2004.07.012
- Wang, X., Mohamed, I., Ali, M., Abbas, M. H. H., Shah, G. M., and Chen, F. (2019). Potassium distribution in root and non-root zones of two cotton genotypes and its accumulation in their organs as affected by drought and potassium stress conditions. *J. Plant Nutr. Soil Sci.* 182, 72–81. doi:10.1002/jpln.201800026
- Wei, H., Chen, S., Niyitanga, S., Liu, T., Qi, J., and Zhang, L. (2022a). Genome-wide identification and expression analysis response to GA(3) stresses of WRKY gene family in seed hemp (*Cannabis sativa* L.). *Gene* 822, 146290. doi:10.1016/j.gene.2022.146290
- Wei, Y. L., Jin, J. P., Liang, D., Gao, J., Li, J., Xie, Q., et al. (2022b). Genome-wide identification of *Cymbidium sinense* WRKY gene family and the importance of its Group III members in response to abiotic stress. *Front. Plant Sci.* 13, 969010. doi:10.3389/fpls.2022.969010
- Yang, H., Xia, J., Cui, Q., Liu, J., Wei, S., Feng, L., et al. (2021). Effects of different *Tamarix chinensis*-grass patterns on the soil quality of coastal saline soil in the Yellow River Delta, China. *Sci. Total Environ.* 772, 145501. doi:10.1016/j.scitotenv.2021.145501
- Yang, Y., Zhao, L., Yang, G., Zang, Y., Fu, P., Hu, J., et al. (2022). Selection and validation of reference genes for leaf color phenotype in 'maiyoujinjiqu', a *Catalpa fargesii* variety, by qRT-PCR. *For. Res.* 35 (1), 123–131. doi:10.13275/j.cnki.lykxyj.2022.01.014
- Yao, W., Wang, S., Zhou, B., and Jiang, T. (2016). Transgenic poplar overexpressing the endogenous transcription factor ERF76 gene improves salinity tolerance. *Tree Physiol.* 36, 896–908. doi:10.1093/treephys/tpw004
- Zhu, H., Jiang, Y., Guo, Y., Huang, J., Zhou, M., Tang, Y., et al. (2021). A novel salt inducible WRKY transcription factor gene, *AhWRKY75*, confers salt tolerance in transgenic peanut. *Plant Physiol. Biochem.* 160, 175–183. doi:10.1016/j.plaphy.2021.01.014
- Zhu, J. (2016). Abiotic stress signaling and responses in plants. *Cell* 167, 313–324. doi:10.1016/j.cell.2016.08.029
- Zou, Z., Yang, L., Wang, D., Huang, Q., Mo, Y., and Xie, G. (2016). Gene structures, evolution and transcriptional profiling of the WRKY gene family in castor bean (*Ricinus communis* L.). *PLoS One* 11, e0148243. doi:10.1371/journal.pone.0148243



OPEN ACCESS

EDITED BY

Krishnanand P. Kulkarni,
Delaware State University, United States

REVIEWED BY

Satinder Kaur,
Punjab Agricultural University, India
Jindong Liu,
Chinese Academy of Agricultural
Sciences, China

*CORRESPONDENCE

Alma Kokhmetova,
✉ gen_kalma@mail.ru
Deepmala Sehgal,
✉ Deepmala.Sehgal@syngenta.com
Nagenahalli Dharmegowda Rathana,
✉ rathanndnagl27@gmail.com
Gopalareddy Krishnappa,
✉ gopalgpb@gmail.com

RECEIVED 24 July 2023

ACCEPTED 27 October 2023

PUBLISHED 23 November 2023

CITATION

Kokhmetova A, Rathana ND, Sehgal D,
Malysheva A, Kumarbayeva M,
Nurzuma M, Bolatbekova A,
Krishnappa G, Gulyaeva E,
Kokhmetova A, Keishilov Z and
Bakhtuly K (2023), QTL mapping for
seedling and adult plant resistance to
stripe and leaf rust in two winter
wheat populations.
Front. Genet. 14:1265859.
doi: 10.3389/fgene.2023.1265859

COPYRIGHT

© 2023 Kokhmetova, Rathana, Sehgal,
Malysheva, Kumarbayeva, Nurzuma,
Bolatbekova, Krishnappa, Gulyaeva,
Kokhmetova, Keishilov and Bakhtuly.
This is an open-access article distributed
under the terms of the [Creative
Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/).
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.

QTL mapping for seedling and adult plant resistance to stripe and leaf rust in two winter wheat populations

Alma Kokhmetova^{1*}, Nagenahalli Dharmegowda Rathana^{2*},
Deepmala Sehgal^{3*}, Angelina Malysheva¹, Madina Kumarbayeva¹,
Makpal Nurzuma¹, Ardak Bolatbekova¹,
Gopalareddy Krishnappa^{4*}, Elena Gulyaeva⁵, Asia Kokhmetova¹,
Zhenis Keishilov¹ and Kanat Bakhtuly¹

¹Laboratory of Breeding and Genetics, Institute of Plant Biology and Biotechnology (IPBB), Almaty, Kazakhstan, ²Coverta Agriscience, Hyderabad, Telangana, India, ³Syngenta, Jealott's Hill International Research Centre, Bracknell, United Kingdom, ⁴ICAR-Sugarcane Breeding Institute, Coimbatore, India, ⁵Laboratory of Mycology and Phytopathology, All Russian Institute of Plant Protection, Pushkin, Russia

The two recombinant inbred line (RIL) populations developed by crossing Almaly × Avocet S (206 RILs) and Almaly × Anza (162 RILs) were used to detect the novel genomic regions associated with adult plant resistance (APR) and seedling or all-stage resistance (ASR) to yellow rust (YR) and leaf rust (LR). The quantitative trait loci (QTLs) were detected through multi-year phenotypic evaluations (2018–2020) and using high-throughput DArTseq genotyping technology. RILs exhibited significant genetic variation with $p < 0.001$, and the coefficient of variation ranged from 9.79% to 47.99% for both LR and YR in all Environments and stages of evaluations. The heritability is quite high and ranged between 0.47 and 0.98. We identified nine stable QTLs for YR APR on chromosomes 1B, 2A, 2B, 3D, and 4D and four stable QTLs for LR APR on chromosomes 2B, 3B, 4A, and 5A. Furthermore, *in silico* analysis revealed that the key putative candidate genes such as *cytochrome P450*, *protein kinase-like domain superfamily*, *zinc-binding ribosomal protein*, *SANT/Myb domain*, *WRKY transcription factor*, *nucleotide sugar transporter*, and *NAC domain superfamily* were in the QTL regions and probably involved in the regulation of host response toward pathogen infection. The stable QTLs identified in this study are useful for developing rust-resistant varieties through marker-assisted selection (MAS).

KEYWORDS

wheat, QTL, yellow rust, leaf rust, adult plant resistance, all-stage resistance

Introduction

Globally, stripe rust or yellow rust (YR) and brown rust or leaf rust (LR) are two important biotic stresses of wheat (*Triticum aestivum* L.). The YR caused by *Puccinia striiformis* (*Pst*) generally causes crop damage in the range of 0.1%–5.0%; however, crop losses can increase to 5%–25% (Wellings, 2011) based on varietal reaction and prevailing environmental conditions, and in severe conditions, crop damage can reach up to 100% (Ali et al., 2014). Most wheat-cultivating areas covering the United States, Eastern and Southern Asia, East Africa, Oceania, the Arabian Peninsula, and Western Europe are vulnerable to *Pst*

incidence. The monoculture of single or closely related cultivars coupled with favorable environmental conditions is ideal for pathogen evolution. Several incidences of YR epidemics have occurred in different parts of Central Asia and Kazakhstan (Yessenbekova et al., 2016; Kokhmetova et al., 2018; Kokhmetova et al., 2021a). The YR incidence in Central and Western Asia has substantially increased between 2001 and 2010 (Morgounov et al., 2013). Recently, Central Asia recorded four YR epidemics between the years 2009 and 2014 (Ziyaev et al., 2011; Sharma et al., 2014). Historically, YR used to be confined to cool weather conditions; however, it has slowly moved to non-conventional regions due to race evolution (Muleta et al., 2017; Godoy et al., 2018). The LR caused by *Puccinia triticina* (Pt) is comparatively less devastating than the other two wheat rusts; however, it causes more crop damage as the frequency of its occurrence is very high and has wide global distribution (Huerta-Espino et al., 2011). The wide adaptability of this rust makes it spread to temperate areas, resulting in approximately 70% of yield losses (Herrera-Foessel et al., 2006; Aktar-Uz-Zaman et al., 2017). North Kazakhstan has reported five leaf rust epidemics during 2001–2009, causing yield damage in the range of 10%–50%, particularly in the susceptible varieties (Kokhmetova et al., 2016; Kokhmetova et al., 2021b).

Rust resistance breeding provides a sustainable solution to protect the wheat from loss of yield and grain quality. Genetically, there are two kinds of rust resistance; one is race-specific seedling resistance or all-stage resistance (ASR) and the other is race-non-specific adult plant resistance (APR) or partial resistance (Chen, 2013). Race-specific resistance is qualitative in nature, is governed by a single gene or oligogenes and only effective against a single or few races, and follows the gene-for-gene hypothesis (Flor, 1971). They express from the seedling to adult plant stage and confer vertical resistance. Generally, race-specific seedling resistance is less durable, which can easily be overcome by race evolution (Jones and Dangl, 2006). In contrast, race-non-specific resistance genes are more durable, but when used alone, they are unable to provide high levels of resistance; however, when used in combination with other race-specific or race-non-specific genes, they provide adequate resistance (Singh et al., 2000).

At present, there are 86 YR genes that were cataloged (McIntosh et al., 2020; Zhu et al., 2023); however, only a few genes like *Yr5* and *Yr15* are effective to most of the prevailing *Pst* races across the globe (Sharma-Poudyal et al., 2013). The Yr gene diversity in commercial cultivars is very important in managing stripe rust epidemics. Additionally, non-race-specific resistance driven by a few Yr genes such as *Yr18*, which express at the adult plant stage, confers field resistance against the three wheat rusts, which have been widely used for several decades (Randhawa et al., 2012; Krattinger et al., 2016). However, single-gene-based resistance in varieties is not enough to protect the cultivars, particularly under high disease pressure conditions (Zhang et al., 2019). Thus, the combined use of APR genes along with one or a few ASR genes may be ideal to protect the cultivars with durable resistance (Ellis et al., 2014; Liu et al., 2018). Similarly, 83 *Lr* genes have been identified (McIntosh et al., 2020; Kolmer et al., 2023), and 15 *Lr* genes exhibited APR response, including *Lr34*, *Lr46*, *Lr67*, *Lr68*, *Lr74*, *Lr75*, *Lr77*, and *Lr78*. Among them, seven are race-specific and eight are race-non-specific (McIntosh et al., 2016). Among race-specific APR genes, *Lr12*, *Lr13*, *Lr22b*, *Lr35*, and *Lr37* are qualitative in

nature and provide hypersensitive reactions only at the adult plant stage (McIntosh et al., 1995; Singh and Bowden, 2011). Previous reports in Kazakhstan revealed that several *Lr* genes became ineffective due to pathogen evolution, resulting in new virulent races. Several *Lr* genes including *Lr9*, *Lr10*, *Lr19*, *Lr34*, *Lr37*, and *Lr68* are still providing resistance to several races, whereas *Lr1* has lost its effectiveness (Koishybaev et al., 2010). Some of the APR genes like *Lr1*, *Lr10*, *Lr21*, *Lr22a*, *Lr34*, and *Lr67* have been cloned; a few cloned genes like *Lr34* and *Lr67* were found to be associated with complex loci conferring resistance to multiple biotic stresses. Few pleiotropic gene complexes like *Lr19/Sr25*, *Lr26/Yr9/Sr31/Pm8*, *Lr37/Yr17/Sr38*, *Lr67/Sr55/Yr46/Pm46*, and *Lr34/Yr18/Pm38/Sr57* are widely used in the breeding programs across the globe, including Kazakhstan, that are still providing sufficient resistance (Kokhmetova et al., 2016; Kokhmetova et al., 2021b).

Although several race-specific seedling genes were identified for YR and LR, the genetic dissection of rust resistance in wheat through QTL mapping is equally important in the management of wheat rust as the durability of most of the race-specific seedling genes is very less, particularly under high disease pressure conditions in regions with a wide distribution of single or similar varieties. The evolution of novel races and the breakdown of race-specific genes led wheat breeders toward identifying and utilizing the durable race-non-specific APR genes and QTLs. The recent advancements in next-generation sequencing (NGS) technologies, the development of the wheat reference genome (IWGSC, 2018), and the cost reduction of genotyping made the genetic dissection of QTL regions and candidate genes more precise and effective. Previously, various mapping populations and marker systems were used to locate QTLs for LR resistance (Kolmer, 2015; Li et al., 2016; Kthiri et al., 2019; Zhang et al., 2019; Bokore et al., 2020; Ciechanowska et al., 2022; Delfan et al., 2023). Similarly, several QTLs were identified for SR resistance in different genetic backgrounds (Wang et al., 2015; Zhang et al., 2019; Farzand et al., 2021; Rollar et al., 2021; Yuan et al., 2020; Cheng et al., 2022; Rauf et al., 2022; Tehseen et al., 2022). However, very few are effective in providing resistance, and many among them provide race-specific resistance and hence have limited applicability to wide area deployment.

Therefore, we designed our study to identify the genomic regions that confer ASR and APR resistance to leaf and stripe rust resistance to the races prevalent across Central Asia, particularly in Kazakhstan, using two RIL mapping populations derived from Almaly × Anza and Almaly × Avocet S with multi-environment evaluations. We also attempted to provide the putative candidate genes for the identified stable QTLs to assist further validation and gene cloning experiments.

Materials and methods

Plant material and field experiments

The parental genotypes used to develop RILs are contrasting for both YR and LR; Almaly was the resistant parent, whereas Anza and Avocet were the susceptible parents. The RIL populations were developed by crossing Almaly × Anza (160 RILs) and Almaly × Avocet S (206 RILs) through the single-seed descent method in

southeastern Kazakhstan (Supplementary Table S1). Since Almaly is the common parent in both the crosses, hereafter Almaly \times Anza will be referred to as the Anza population, whereas Almaly \times Avocet S will be referred to as the Avocet population. The RILs were evaluated at the seedling and adult plant-growth stages for LR and YR pathogens. The mapping populations of both the crosses along with parental genotypes were evaluated at the Kazakh Research Institute of Agriculture and Crop Production (KazNIIZiR), Almalybak (43°13'09"N and 76°36'17"E) for 2 consecutive years during 2018–19 and 2019–20 for YR and LR APR, respectively. Additionally, the RIL population of Almaly \times Anza was evaluated during 2020–21 for LR APR in a randomized complete block design (RCBD) following the two replications. Each RIL was sown in a two-row plot of 1.5 m length and row-to-row spacing of 25 cm. The susceptible check variety, Morocco, was planted at an interval of every 20 plots. The RILs were also evaluated for YR and LR ASR in a greenhouse facility at the All-Russian Institute of Plant Protection (ARIPP), St Petersburg, Russia (59°73'73"N, 30°42'47"E) during 2020. Three to five seeds of each genotype were planted in 10-cm-diameter plastic pots in a disease-free area. The RILs were inoculated after 7–10 days under greenhouse conditions with three races of *P. striiformis* and six races of *P. recondita* with different levels of virulence to *Lr* and *Yr* genes (Supplementary Table S3). All entries were arranged in an RCBD design with three replications. The complete phenotypic data file of two biparental populations is provided in Supplementary Table S1.

Phenotyping

Seedling resistance in greenhouse

The *P. striiformis* races were differentiated in 2020 using a set of 12 wheat lines developed in the Avocet wheat background and on nine supplemental wheat differential lines using a method developed by Johnson et al. (1972). The determination of the type of plant reaction was carried out twice within 14–20 days after infection according to the Gassner and Straib accounting scale (Gassner and Straib, 1932). At the same time, the reactions of 0, 1, and 2 points were assigned to the resistant type R (Resistant), and those of 3 and 4 points were assigned to the susceptible type S (Susceptible). The *P. triticea* races were also differentiated during 2020 using 20 near-isogenic lines (NILs) developed in the Thatcher background, each carrying one of the LR resistant genes (Kolmer and Ordenez, 2007; Schachtel et al., 2012; Kolmer et al., 2014). The virulence of the phenotypes was determined on these 20 differential lines and encoded with 0 and 1 for avirulence and virulence, respectively (Long and Kolmer, 1989; Kolmer and Ordenez, 2007). The virulence analysis tools (Schachtel et al., 2012) was used for the nomenclature of *P. triticea* races. The type of response to leaf rust was determined twice within 14–20 days after infection, according to the scale of Mains and Jackson (1926). The reactions of 0, 1, and 2 points were assigned to the resistant type R (Resistant), and those of 3 and 4 points were assigned to the susceptible type S (Susceptible).

The seedlings of the RIL population from Almaly \times Avocet S cross along with the parents were inoculated with two races of *P. striiformis*, i.e., 108E187 (Pst_1) and 110E191 (Pst_2), and two races of *P. triticea*, i.e., MLTTH and TLTTT, to determine the race-specific resistance. Similarly, the seedlings of the RIL population

from Almaly \times Anza cross along with parents were inoculated with two races of *P. striiformis*, i.e., 108E187 (Pst_1) and 101E191 (Pst_3), and four races of *P. triticea*, i.e., THTTQ, TCTTR, TCPTQ, and THTTT. The plants were infected with spores at a three-leaf stage, and plants were placed in a humid chamber for 24 h. The seedling infection type of the RIL was scored using the same approach as that for races differentiation. The pathotypes used in this study and their virulence reaction to rust genes are provided in Supplementary Table S3.

Phenotyping for adult plant resistance in the field

The field phenotyping for YR and LR APR was carried out during 2018–2019 for both the populations and also during 2020 for LR APR for the Anza population at Kazakh Research Institute of Agriculture and Crop Production (KazNIIZiR), Almalybak. Pathogen racial mixtures from the local population were used to inoculate the mapping populations. The method proposed by Roelfs et al. (1992) was followed for spore sampling, storage, and propagation. The pathogen was propagated in a greenhouse on the susceptible wheat variety, Morocco. The experimental wheat material was inoculated with a mixture of spores and talc in the ratio of 1:100 by spraying with an aqueous suspension of spores with 0.001% Tween-80 at the stem elongation stages (Z21–32). After infection, the plots were wrapped with a plastic cover for 16–18 h to create high humidity. After the manifestation of diseases on susceptible control varieties, an assessment (2–3 times) of rust resistance was carried out. Leaf and yellow rust resistance of wheat accessions was evaluated using the modified Cobb scale (Peterson et al., 1948; McIntosh et al., 1995). The scoring was based both on disease severity (proportion of the leaf area infected) and on the plant response to infection (reaction type). Plant responses were recorded as resistant (R), moderately resistant (MR), moderately susceptible (MS), and susceptible (S) reactions.

Phenotypic analysis

The phenotypic analysis was done in multi environment trial analysis with R (META-R) version 6.0 software. The best linear unbiased predictors (BLUPs) for each year and across year were used for QTL analysis. Furthermore, genetic correlation among traits and between environments, heritability, and ANOVA was done using META-R. The details of the analysis are provided in Rathan et al. (2023). Past V 3.01 was used to generate frequency distribution graphs.

Genotyping

The genomic DNA was extracted from the parents, and each RIL was extracted from both the populations following the modified cetyltrimethylammonium bromide (CTAB) method (Dreisigacker et al., 2012). The DArTseq technology was used for genotyping of both the RILs in the Genetic Analysis and Service for Agriculture (SAGA) lab based in Mexico (Edet et al., 2018). Briefly, the sequencing of mapping populations was carried out at 192-plexing on Illumina HiSeq2500 with 1 \times 77-bp reads. Allele calls

TABLE 1 Genetic parameters of Almaly × Anza and Almaly × Avocet RIL populations.

Year	Trait	Heritability	Genotype variance	Gen × Envi variance	Grand mean	LSD	CV (%)
Almaly × Anza RIL population							
2018	LR_APR	0.97	909.10***	—	24.30	15.23	32.27
	YR_APR	0.96	906.32***	—	32.04	16.49	26.58
2019	LR_APR	0.88	231.48***	—	16.54	14.71	47.99
	YR_APR	0.90	331.35***	—	20.11	15.82	41.93
2020	LR_APR	0.88	162.38***	—	10.52	12.33	63.30
	LR_ASR_THTTQ	0.82	1.12***	—	2.45	1.24	28.37
	LR_ASR_TCTTR	0.92	1.89***	—	2.46	1.11	24.01
	LR_ASR_TCPTQ	0.93	2.22***	—	2.04	1.07	27.52
	LR_ASR_THTTTR	0.88	1.41***	—	1.62	1.17	38.85
	YR_ASR_Pst1	0.97	1.65***	—	2.68	0.67	12.90
	YR_ASR_Pst3	0.94	2.06***	—	1.57	0.98	32.67
Overall	LR_APR	0.87	315.88***	118.43***	17.12	18.29	43.82
	YR_APR	0.71	362.58***	256.25***	26.07	28.61	32.50
Almaly × Avocet RIL population							
2018	LR_APR	0.95	791.09***	—	32.64	17.27	27.52
	YR_APR	0.93	687.52***	—	48.67	19.64	21.25
2019	LR_APR	0.89	288.97***	—	22.01	15.49	37.77
	YR_APR	0.95	1,095.34***	—	56.96	20.49	18.71
2020	LR_ASR_MLTTH	0.94	2.21***	—	1.67	0.97	30.38
	LR_ASR_TLTTR	0.90	1.43***	—	1.29	1.05	43.49
	YR_ASR_Pst1	0.96	1.42***	—	3.43	0.65	9.79
	YR_ASR_Pst2	0.98	3.37***	—	1.77	0.71	20.56
Overall	LR_APR	0.47	177.93***	362.09***	27.33	27.09	31.67
	YR_APR	0.79	620.46***	270.97***	52.81	31.82	19.88

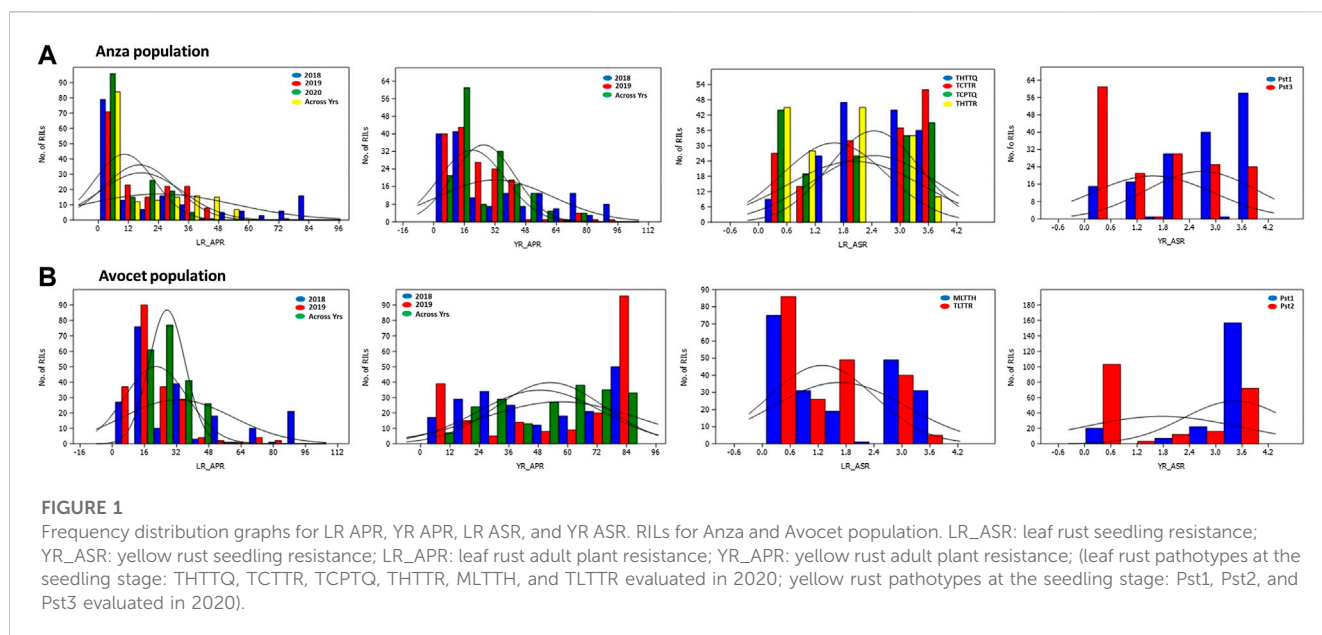
LSD, least significant difference, CV, coefficient of variation, LR_ASR, leaf rust seedling resistance, YR_ASR, yellow rust seedling resistance; LR_APR, leaf rust adult plant resistance; YR_APR, yellow rust adult plant resistance; LR_APR, leaf rust adult plant resistance (leaf rust pathotypes at the seedling stage, THTTQ, TCTTR, TCPTQ, THTTTR, MLTTH, and TLTTR; yellow rust pathotypes at the seedling stage, Pst1, Pst2, and Pst3); *** significance at $p < 0.001$; ** significance at $p < 0.01$.

for SNPs were generated through the proprietary analytical pipeline developed by DArT P/L (Sansaloni et al., 2011). Furthermore, the genetic locations of the SNPs were identified by using a 100 K consensus map given by SAGA (Sansaloni et al. unpublished). The complete genotypic data for the two biparental populations are provided in [Supplementary Table S2](#).

Linkage mapping and QTL detection

The linkage maps were constructed separately for Anza and Avocet RIL populations using DArTseq SNP markers. The procedure followed for linkage map construction and QTL detection is the same for both populations. The markers were filtered, and the monomorphic markers, markers with >30% missing data, high heterozygosity percentage (>30%) and low

allele frequency (<5%) were removed. The BIN functionality in IciMapping 4.2 QTL software was used to remove redundant markers. A filtered set of 1,293 and 1,127 high-quality SNPs were finally used for QTL analysis in Anza and Avocet populations, respectively. The linkage map construction and QTL mapping was done in IciMapping 4.2 QTL software (Wang et al., 2012; Li et al., 2016). The Kosambi mapping function was used to construct linkage groups, using a threshold logarithm of odds (LOD) score of 3.0 (Kosambi, 1943). Within each linkage group, the marker order was carried out with the 2-opt algorithm, and rippling was carried out by maintaining a window size of 5 cM. QTL mapping was done using complex composite interval additive functionality mapping (ICIM-ADD) (Li et al., 2008). Additive QTLs were detected using a 1.0 cM incremental scan. The LOD log confidence for QTL mapping was chosen as 3.0. Then, the QTLs were localized on



the respective chromosomes. One-LOD drop from the estimated QTL position was considered the confidence interval.

In silico analysis

Stable QTLs with high phenotypic variation were used for the identification of candidate genes. The genes were identified in the RefSeq v1.0 assembly from the International Wheat Genome Sequencing Consortium (IWGSC) integrated in the Ensembl Plant database (<https://plants.ensembl.org/index.html>) using the basic local alignment search tool (BLAST). The molecular functions of the probable candidate genes found in the overlapping regions and within the 0.1 Mb flanking regions were identified. The role of the genes in governing leaf and yellow rust resistance was validated by comparing with the published literatures.

Results

Genetic parameters and trait associations

Genetic parameters of both the RIL populations derived from Anza and Avocet crosses are presented in Table 1. Wide variability exists for both YR and LR resistance in both the RIL populations for all the races, as evidenced by the presence of a highly significant genotypic variance (Table 1). The frequency distributions of YR and LR severity in the field for RILs from both populations and in the seedling stages exhibited continuous variation (Figure 1; Supplementary Figures S1, S2). The interaction between genotype and location was significant for the pooled mean of YR and LR APR for both the RIL populations. Both populations exhibited a high broad sense heritability (≥ 0.8) for all the traits, except the APR of yellow rust (0.71) pooled data in the Anza population and APR of leaf rust (0.47) and yellow rust (0.79) pooled data in Avocet RIL

population. The CV ranged from 12.9% (YR_ASR_Pst1 in 2020) to 63.3% (LR_APR in 2020) for the Anza population. Similarly, the CV ranged from 9.79% (YR_ASR_Pst1 in 2020) to 43.49% (LR_ASR_TLTTR in 2020) for the Avocet population. Broad sense heritability estimates (h^2) for leaf and yellow rust across years and different infection backgrounds were high (0.82–0.98), indicating that rust resistance can be improved by breeding (Table 1).

The year-wise genetic correlations between YR APR and LR APR in Anza and Avocet populations are presented in Table 2. A significant correlation was found between YR and LR APR during 2019 ($p < 0.05$) and the overall mean ($p < 0.01$) in Anza population; however, no correlation was observed in 2018. Furthermore, in the Avocet population, no correlation was observed between the traits.

Marker statistics

Genotyping of both Anza and Avocet populations was carried out using next-generation sequencing technology DArTseq™ (<http://www.diversityarrays.com/dart-application-dartseq>). A filtered set of 1,293 and 1,127 high-quality SNPs were used for linkage map construction and QTL identification in Anza and Avocet populations, respectively (Table 3). In Anza population, 539 SNPs were mapped on A subgenome, 491 SNPs on B subgenome, and only 263 SNPs on D subgenome in Anza population, whereas in Avocet cross, 482 SNPs were mapped on the B subgenome, 423 SNPs on A subgenome, and 222 SNPs on D subgenome.

QTL analysis

The QTLs identified for APR and ASR for both the rusts are presented in Tables 4, 5 and illustrated in Figures 2, 3. A set of 51 QTLs were identified, out of which 28 QTLs included LR APR (6 QTLs), YR APR (10), and six each for LR and YR ASR in the Anza

TABLE 2 Genetic correlations of Almaly × Anza and Almaly × Avocet RIL populations.

Almaly × Anza RIL population		
Year	Trait	LR_APR
2018	YR_APR	0.13
2019	YR_APR	0.17*
Overall	YR_APR	0.43**
Almaly × Avocet RIL population		
Year	Trait	YR_APR
2018	LR_APR	0.01
2019	LR_APR	0.1
Overall	LR_APR	0.04

LR_APR, leaf rust adult plant resistance; YR_APR, yellow rust adult plant resistance; ** significance at $p < 0.01$; * significance at $p < 0.05$.

TABLE 3 Marker distribution in Almaly × Anza and Almaly × Avocet RIL populations.

Chromosome/subgenome	1	2	3	4	5	6	7	Total
Almaly × Anza RIL population								
A subgenome	66	61	103	80	80	58	91	539
B subgenome	85	71	86	55	79	47	68	491
D subgenome	49	42	33	25	30	45	39	263
Almaly × Avocet RIL population								
A subgenome	59	59	62	78	78	9	78	423
B subgenome	91	72	73	40	74	65	67	482
D subgenome	36	33	22	21	29	36	45	222

population. The remaining 23 QTLs were identified in the Avocet population including LR APR (3 QTLs), YR APR (12 QTLs), LR ASR (5 QTLs), and YR ASR (3 QTLs). Subgenome-wise 17 QTLs were identified on each subgenome of A, B, and D considering both the populations. Furthermore, the information about the favorable alleles of consistent QTLs is provided in [Supplementary Table S5](#).

Anza population

QTLs for yellow and leaf rust APR

For LR APR, a total of six QTLs, i.e., *QLR-APR-4A*, *QLR-APR-5B*, *QLR-APR-1A*, *QLR-APR-2B*, *QLR-APR-3B*, and *QLR-APR-7D*, were identified on different chromosomes at 233 cM, 288 cM, 97 cM, 295–313 cM, 340–342 cM and, 211 cM, respectively. The identified QTLs explained that the percent phenotypic variation ranged from 1.16 (*QLR-APR-3B*) to 15.18 (*QLR-APR-2B*). A maximum of three QTLs was identified on B, two QTLs on A, and one QTL on D subgenomes ([Table 4](#)).

For YR APR, 10 QTLs were mapped on 1D, 2A, 4D, 7D, 3A, and 3D at different locations. The 10 identified QTLs, i.e., *QYR-APR-2A.1*, *QYR-APR-2A.2*, *QYR-APR-1D*, *QYR-APR-4D.1*, *QYR-APR-7D*, *QYR-APR-3A.3*, *QYR-APR-4D.2*, *QYR-APR-3A.1*, *QYR-APR-*

3A.2, and *QYR-APR-3D*, explained the percent phenotypic variation of 9.94, 10.39, 4.74, 8.88, 7.34, 10.57, 9.36, 10.35, 20.93, and 4.26, respectively. A maximum of five QTLs was observed, each on A and D subgenomes; however, no QTL was identified on the B subgenome.

Stable QTLs for APR

A total of six consistent QTLs including three each for LR APR (*QLR-APR-4A*, *QLR-APR-2B*, and *QLR-APR-3B*) and YR APR (*QYR-APR-2A.1*, *QYR-APR-2A.2*, and *QYR-APR-4D.2*) were detected in the Anza population. One QTL, i.e., *QLR-APR-2B*, was detected during 2019 and 2020, which were flanked between marker intervals 3064426–1125988 at a confidence interval of 287.5 cM–322.5 cM with 15.18% (2019) and 9.41% (2020) PVE, respectively ([Table 4](#)). The remaining two consistent QTLs, i.e., *QLR-APR-4A* and *QLR-APR-3B*, were identified in one environment and pooled means. *QLR-APR-4A* was flanked between marker intervals 998,585–1054130 and confidence intervals 230.5 cM–233.5 cM with 8.08 (2018) and 5.76% PVE, respectively (across years), whereas *QLR-APR-3B* was flanked between marker intervals 2257185–4396068 and confidence intervals 325.5 cM–342 cM with 8.08 (2020) and 5.76% PVE, respectively (across years).

TABLE 4 QTLs identified for yellow and leaf rust resistance in Almaly × Anza RIL population locations for 3 years.

Year	Trait	QTL	Chr	Genetic position (cM)	Physical position (bp)	Flanking markers	LOD	PVE (%)	Add	Confidence interval
2018	LR-APR	<i>QLR-APR-4A</i>	4A	233	652,566,572–716,874,949	998,585–1,054,130	3.35	8.08	9.04	230.5–233.5
	LR-APR	<i>QLR-APR-5B</i>	5B	288	18,616,930–529,978,625	1,246,645–4,990,876	4.01	9.94	−9.72	277.5–288
2019	LR-APR	<i>QLR-APR-1A</i>	1A	97	18,885,429	100,091,105–1,151,033	3.69	6.23	4.17	94.5–98.5
	LR-APR	<i>QLR-APR-2B</i>	2B	295	785,575,596	3,064,426–1,125,988	5.83	15.18	6.39	287.5–314.5
2020	LR-APR	<i>QLR-APR-2B</i>	2B	313	785,575,596	3,064,426–1,125,988	6.23	9.41	9.36	305.5–322.5
	LR-APR	<i>QLR-APR-3B</i>	3B	342	27,980,930–457,925,422	2,257,185–4,396,068	3.03	1.16	3.19	325.5–342
	LR-APR	<i>QLR-APR-7D</i>	7D	211	42,075,096–670,266,622	1,126,655–1,214,912	3.42	4.81	6.66	200.5–222.5
Across years	LR-APR	<i>QLR-APR-4A</i>	4A	233	652,566,572 –716,874,949	998,585–1,054,130	4.34	5.76	5.39	231.5–233.5
	LR-APR	<i>QLR-APR-3B</i>	3B	340	27,980,930–457,925,422	2,257,185–4,396,068	4.16	6.37	5.51	326.5–342
2018	YR-APR	<i>QYR-APR-2A.1</i>	2A	102	683,536,393–709,771,711	2,260,254–3,064,488	6.15	9.94	10.16	96.5–106.5
	YR-APR	<i>QYR-APR-2A.2</i>	2A	174	20,557,628–34,846,399	1,212,067–1,242,826	6.86	10.39	10.40	169.5–180.5
	YR-APR	<i>QYR-APR-1D</i>	1D	129	378,730,760	1,100,394–100,081,053	3.36	4.74	−7.40	124.5–133.5
	YR-APR	<i>QYR-APR-4D.1</i>	4D	81	483,069,995–497,452,565	1,214,617–4,910,613	4.86	8.88	−9.60	77.5–87.5
	YR-APR	<i>QYR-APR-7D</i>	7D	96	637,319,694	1,158,021–4,022,626	4.37	7.34	−8.74	86.5–99.5
2019	YR-APR	<i>QYR-APR-3A.3</i>	3A	179	60,385,881	100,080,358–1,045,110	4.48	10.57	−5.31	178.5–179.5
	YR-APR	<i>QYR-APR-4D.2</i>	4D	106	495,101,244–498,684,043	1,133,723–1,012,563	3.94	9.36	−4.99	100.5–110.5
Across years	YR-APR	<i>QYR-APR-2A.1</i>	2A	106	683,536,393–709,771,711	2,260,254–3,064,488	4.95	4.07	4.65	101.5–110.5
	YR-APR	<i>QYR-APR-2A.2</i>	2A	175	9,619,406–20,557,628	1,242,826–1,230,957	6.16	4.43	4.84	169.5–182.5
	YR-APR	<i>QYR-APR-3A.1</i>	3A	159	590,301,966–603,234,500	1,068,094–1,150,748	12.67	10.35	7.65	158.5–159.5
	YR-APR	<i>QYR-APR-3A.2</i>	3A	164	567,971,052–570,530,990	1,090,173–1,083,292	23.19	20.93	−10.67	162.5–164.5
	YR-APR	<i>QYR-APR-3D</i>	3D	266	589,118,333–603,727,504	1,128,362–1,091,629	5.56	4.26	4.80	260.5–268.5
	YR-APR	<i>QYR-APR-4D.2</i>	4D	107	498,684,043–500,497,895	1,012,563–3,936,672	11.15	8.56	−6.73	104.5–110.5
2020	LR-ASR-THTTTR	<i>QLR-ASR-THTTTR-7A</i>	7A	144	701,561,539–713,432,629	1,111,941–1,125,395	4.31	9.07	0.37	137.5–144.5
	LR-ASR-THTTTR	<i>QLR-ASR-THTTTR-3B</i>	3B	131	736,747,190–754,143,752	1,076,415–3,064,587	3.14	6.71	−0.32	126.5–133.5

(Continued on following page)

TABLE 4 (Continued) QTLs identified for yellow and leaf rust resistance in Almaly × Anza RIL population locations for 3 years.

Year	Trait	QTL	Chr	Genetic position (cM)	Physical position (bp)	Flanking markers	LOD	PVE (%)	Add	Confidence interval
	LR-ASR-TCTTR	<i>QLR-ASR-TCTTR-6B.1</i>	6B	32	11,050,813–74,488,850	4,988,974–1,128,034	3.17	5.07	1.03	25.5–38.5
	LR-ASR-TCTTR	<i>QLR-ASR-TCTTR-6B.2</i>	6B	79	NA	7,353,355–100,069,075	3.53	5.14	1.05	74.5–84.5
	LR-ASR-TCPTQ	<i>QLR-ASR-TCPTQ-2D</i>	2D	4	623,163,577–641,940,538	2,256,914–2,250,689	5.02	3.70	0.52	0–9.5
	LR-ASR-TCPTQ	<i>QLR-ASR-TCPTQ-6D</i>	6D	223	485,975,428	100,023,455–1,091,595	4.15	14.63	1.05	216.5–228.5
2020	YR-ASR-Pst3	<i>QYR-ASR-Pst3-1B</i>	1B	77	630,355,967–679,858,781	1,230,145–1,273,377	5.33	4.12	1.14	71.5–82.5
	YR-ASR-Pst3	<i>QYR-ASR-Pst3-6B.1</i>	6B	28	11,050,813–74,488,850	4,988,974–1,128,034	6.47	4.24	1.16	21.5–35.5
	YR-ASR-Pst3	<i>QYR-ASR-Pst3-6B.2</i>	6B	82	NA	7,353,355–100,069,075	5.80	4.19	1.15	76.5–88.5
	YR-ASR-Pst3	<i>QYR-ASR-Pst3-6B.3</i>	6B	171	668,517,583–691,343,000	1,095,762–1,250,690	3.29	0.71	–0.47	164.5–174.5
	YR-ASR-Pst3	<i>QYR-ASR-Pst3-7B</i>	7B	234	167,570,301	100,078,188–1,025,576	7.38	4.25	–1.16	224.5–242.5
	YR-ASR-Pst1	<i>QYR-ASR-Pst1-6D</i>	6D	435	486,282,549–492,098,588	1,083,737–1,068,228	3.17	8.57	0.39	431.5–438.5

LR_AS, leaf rust seedling resistance, YR_AS, yellow rust seedling resistance; LR_APR, leaf rust adult plant resistance; YR_APR, yellow rust adult plant resistance (leaf rust pathotypes at the seedling stage, THHTQ, TCTTR, TCPTQ, THHTR, MLTTH, and TLTTT; yellow rust pathotypes at the seedling stage, Pst1, Pst2, and Pst3), QTL: quantitative trait locus, cM, centimorgan, LOD, logarithm of odds, PVE, phenotypic variation explained, Add, additive effect; consistent QTLs are highlighted in red font; physical position are obtained from the reference genome IWGSC, RefSeq v2.0.

Similarly, three consistent/stable QTLs, i.e., *QYR-APR-2A.1*, *QYR-APR-2A.2*, and *QYR-APR-4D.2*, were identified for yellow rust. *QYR-APR-2A.1* was mapped between marker intervals 2260254–3064488 and confidence intervals 96.5 cM–110.5 cM with 9.94% (2018) and 4.07% PVE, respectively (across years); similarly, *QYR-APR-2A.2* was mapped between marker intervals 1212067–1242826 (2018) and 1242826–1230957 (across years) at a confidence interval of 169.5 cM–182.5 cM with PVE of 10.39% (2018) and 4.43% (across years), respectively. The third consistent QTL (*QYR-APR-4D.2*) was mapped between the flanking markers of 1133723–1012563 (2019) and 1012563–3936672 (across years) at a confidence interval of 100.5 cM–110.5 cM with PVE of 9.36% (2019) and 8.56% (across years), respectively.

QTLs for yellow and leaf rust ASR

Twelve QTLs, including six each for LR (*QLR-ASR-THHTT-7A*, *QLR-ASR-THHTT-3B*, *QLR-ASR-TCTTR-6B.1*, *QLR-ASR-TCTTR-6B.2*, *QLR-ASR-TCPTQ-2D*, and *QLR-ASR-TCPTQ-6D*) and YR (*QYR-ASR-Pst3-1B*, *QYR-ASR-Pst3-6B.1*, *QYR-ASR-Pst3-6B.2*, *QYR-ASR-Pst3-6B.3*, *QYR-ASR-Pst3-7B*, and *QYR-ASR-Pst1-6D*) were identified in the Anza population. For leaf rust, the highest PVE was reported for *QLR-ASR-TCPTQ-6D* (14.63%) followed by *QLR-ASR-THHTT-7A* (9.07%), *QLR-ASR-THHTT-3B* (6.71%), *QLR-ASR-TCTTR-6B.2* (5.14%), *QLR-ASR-TCTTR-6B.1* (5.07%), and *QLR-ASR-TCPTQ-2D* (3.7%). Similarly, for yellow rust, the

highest PVE was reported for *QYR-ASR-Pst1-6D* (8.57%) followed by *QYR-ASR-Pst3-7B* (4.25%), *QYR-ASR-Pst3-6B.1* (4.24%), and *QYR-ASR-Pst3-6B.2* (4.19%).

Avocet population

QTLs for yellow and leaf rust APR

A set of 15 QTLs, including LR (3 QTLs) and YR APR (12 QTLs), were identified in the Avocet population. For leaf rust, three QTLs, i.e., *QLR-APR-5A.1*, *QLR-APR-5A.2*, and *QLR-APR-7A*, flanked between marker intervals 1128503–2262017, 3064895–3958580, and 1102911–1309112, respectively, at a confidence interval of 201.5 cM–202.5 cM, 284.5 cM–312.5 cM, and 86.5 cM–90.5 cM. The highest PVE was reported for *QLR-APR-7A* (11.77%) followed by *QLR-APR-5A.1* (2.79%) and *QLR-APR-5A.2* (2.17%). All the QTLs were mapped only on subgenome A, and there is no representation from subgenomes B and D (Table 5).

For yellow rust APR, 12 QTLs were identified on different chromosomes of all three subgenomes of wheat. The highest PVE was reported for *QYR-APR-4D.2* (16.42% in 2018 and 8.76% across years) followed by *QYR-APR-3D* (12.24% across years and 11.43% in 2018), *QYR-APR-1B* (10.20%), *QYR-APR-4D.1* (8.55%), and *QYR-APR-1A* (7.43%), and the remaining seven QTLs reported less than 7.0% PVE. The highest QTLs were reported in subgenome D followed by subgenomes B and A.

TABLE 5 QTLs identified yellow and leaf rust resistance in Almaly × Avocet RIL population locations for 3 years.

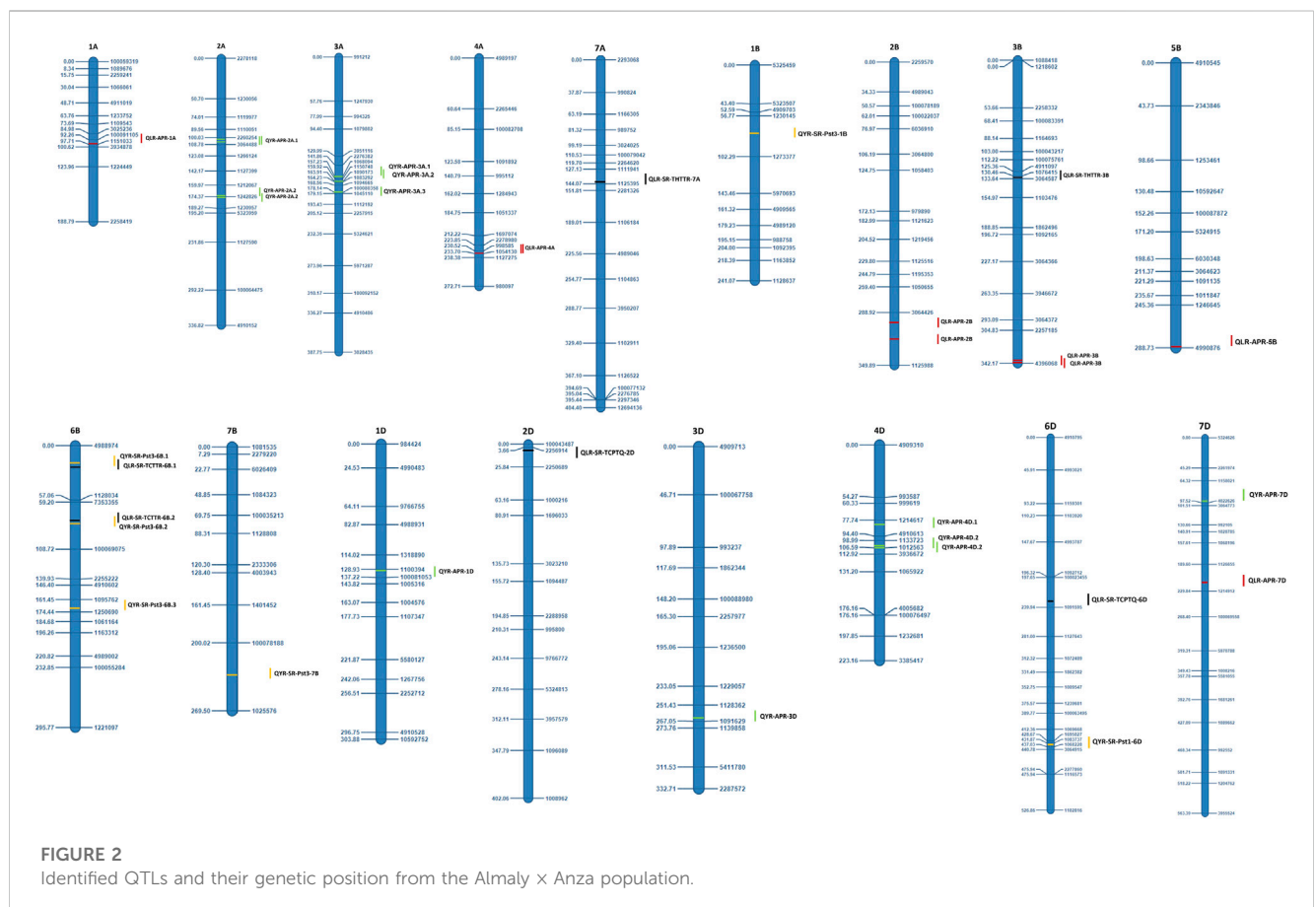
Year	Trait	QTL	Chr	Genetic position (cM)	Physical position (bp)	Flanking markers	LOD	PVE (%)	Add	Confidence interval
2018	LR-APR	QLR-APR-5A.1	5A	202	84,082,776–382,970,680	1,128,503–2,262,017	6.19	2.79	9.87	201.5–202.5
	LR-APR	QLR-APR-5A.2	5A	294	80,081,823–607,673,215	3,064,895–3,958,580	3.86	2.17	–8.93	286.5–297.5
	LR-APR	QLR-APR-7A	7A	88	18,568,134–254,687,847	1,102,911–1,309,112	3.56	11.77	21.99	86.5–90.5
Across years	LR-APR	QLR-APR-5A.2	5A	293	80,081,823–607,673,215	3,064,895–3,958,580	3.63	1.74	–2.89	284.5–312.5
2018	YR-APR	QYR-APR-2A	2A	198	NA	1,106,494–100,058,689	3.52	2.43	5.49	197.5–203.5
	YR-APR	QYR-APR-7A	7A	238	86,051,773–91,538,169	7,350,555–3,064,562	5.58	3.85	6.95	237.5–241.5
	YR-APR	QYR-APR-2B.1	2B	51	7,726,779	6,050,347–1,275,640	4.06	2.85	5.94	48.5–58.5
	YR-APR	QYR-APR-3D	3D	201	612,659,895	3,064,599–100,087,630	10.52	11.43	11.91	193.5–208.5
	YR-APR	QYR-APR-4D.1	4D	169	482,004,603–495,101,244	1,133,723–1,667,202	6.47	6.25	–8.94	163.5–176.5
	YR-APR	QYR-APR-4D.2	4D	248	23,317,222–84,937,400	1,201,923–4,909,310	7.48	16.42	14.26	233.5–255.5
	YR-APR	QYR-APR-5D	5D	6	88,551,618–340,292,671	7,487,719–2,265,426	3.51	3.99	–7.04	0–16.5
2019	YR-APR	QYR-APR-1A	1A	133	17,405,618	991,036–3,934,878	5.71	7.43	9.02	129.5–135.5
	YR-APR	QYR-APR-1B	1B	47	670,142,832–679,858,781	1,230,145–4,005,225	8.18	10.20	10.55	46.5–47.5
	YR-APR	QYR-APR-2B.2	2B	156	220,348,494–235,259,298	5,577,199–1,054,964	4.90	5.93	8.06	155.5–156.5
	YR-APR	QYR-APR-4D.1	4D	165	482,004,603–495,101,244	1,133,723–1,667,202	6.81	8.55	–9.98	163.5–170.5
	YR-APR	QYR-APR-4D.2	4D	231	23,317,222–360,914,321	1,001,325–1,201,923	7.08	8.76	9.79	224.5–246.5
Across years	YR-APR	QYR-APR-1B	1B	47	670,142,832–679,858,781	1,230,145–4,005,225	3.99	4.10	4.75	46.5–47.5
	YR-APR	QYR-APR-2B.1	2B	51	7,726,779	6,050,347–1,275,640	5.94	6.54	6.03	48.5–58.5
	YR-APR	QYR-APR-2B.2	2B	153	192,955,204–193,062,823	2,256,116–1,050,655	6.54	6.91	6.17	152.5–154.5
	YR-APR	QYR-APR-2B.3	2B	241	781,214,059–797,605,252	1,219,456–1,121,623	3.09	5.40	5.47	229.5–249.5
	YR-APR	QYR-APR-1D	1D	50	9,038,548–11,353,637	1,107,347–1,228,408	3.31	3.39	–4.33	35.5–55.5
	YR-APR	QYR-APR-3D	3D	211	612,659,895	3,064,599–100,087,630	3.23	12.24	8.23	201.5–218.5
	YR-APR	QYR-APR-4D.1	4D	165	482,004,603–495,101,244	1,133,723–1,667,202	5.23	5.58	–5.72	163.5–172.5
	YR-APR	QYR-APR-4D.2	4D	240	23,317,222–84,937,400	1,201,923–4,909,310	3.36	6.92	6.18	232.5–255.5

(Continued on following page)

TABLE 5 (Continued) QTLs identified yellow and leaf rust resistance in Almaly \times Avocet RIL population locations for 3 years.

Year	Trait	QTL	Chr	Genetic position (cM)	Physical position (bp)	Flanking markers	LOD	PVE (%)	Add	Confidence interval
2020	LR-ASR-MLTTH	QLR-ASR-MLTTH-2A	2A	307	755,073,581–771,366,169	1,230,056–2,259,439	8.93	6.10	–1.17	299.5–312.5
	LR-ASR-MLTTH	QLR-ASR-MLTTH-5B	5B	127	561,598,758	100,058,433–2,267,368	3.07	4.30	–0.99	126.5–127.5
	LR-ASR-MLTTH	QLR-ASR-MLTTH-6B	6B	23	378,424,485–692,080,108	1,221,097–7,353,355	5.66	5.81	–1.15	19.5–28.5
	LR-ASR-MLTTH	QLR-ASR-MLTTH-3D	3D	105	268,920,926	100,080,288–1,210,613	3.67	0.82	–0.43	97.5–107.5
	LR-ASR-TLTTR	QLR-ASR-TLTTR-6D	6D	213	462,479,335–479,773,419	1,040,130–1,063,571	4.99	7.58	–0.37	209.5–215.5
2020	YR-ASR-Pst2	QYR-ASR-Pst2-3A	3A	141	25,391,100–25,669,110	4,988,975–2,257,915	3.73	4.75	0.48	139.5–141.5
	YR-ASR-Pst2	QYR-ASR-Pst2-5A	5A	93	695,468,777	1,184,257–100,044,187	9.14	14.01	–0.82	89.5–96.5
	YR-ASR-Pst2	QYR-ASR-Pst2-6D	6D	126	25,673,385–51,985,701	1,046,205–1,009,547	3.64	5.80	0.54	120.5–135.5

LR_AS^R, leaf rust seedling resistance; YR_AS^R, yellow rust seedling resistance; LR_AP^R, leaf rust adult plant resistance; YR_AP^R, yellow rust adult plant resistance; (leaf rust pathotypes at the seedling stage, THTTQ, TCTTR, TCPTQ, THTTR, MLTTH, and TLTTR; yellow rust pathotypes at the seedling stage, Ps1, Ps2, and Ps3), QTL, quantitative trait locus, cM, centimorgan, LOD, logarithm of odds, PVE, phenotypic variation explained, Add, additive effect; consistent QTLs are highlighted in red font; physical positions are obtained from the reference genome IWGSC RefSeq v2.0.



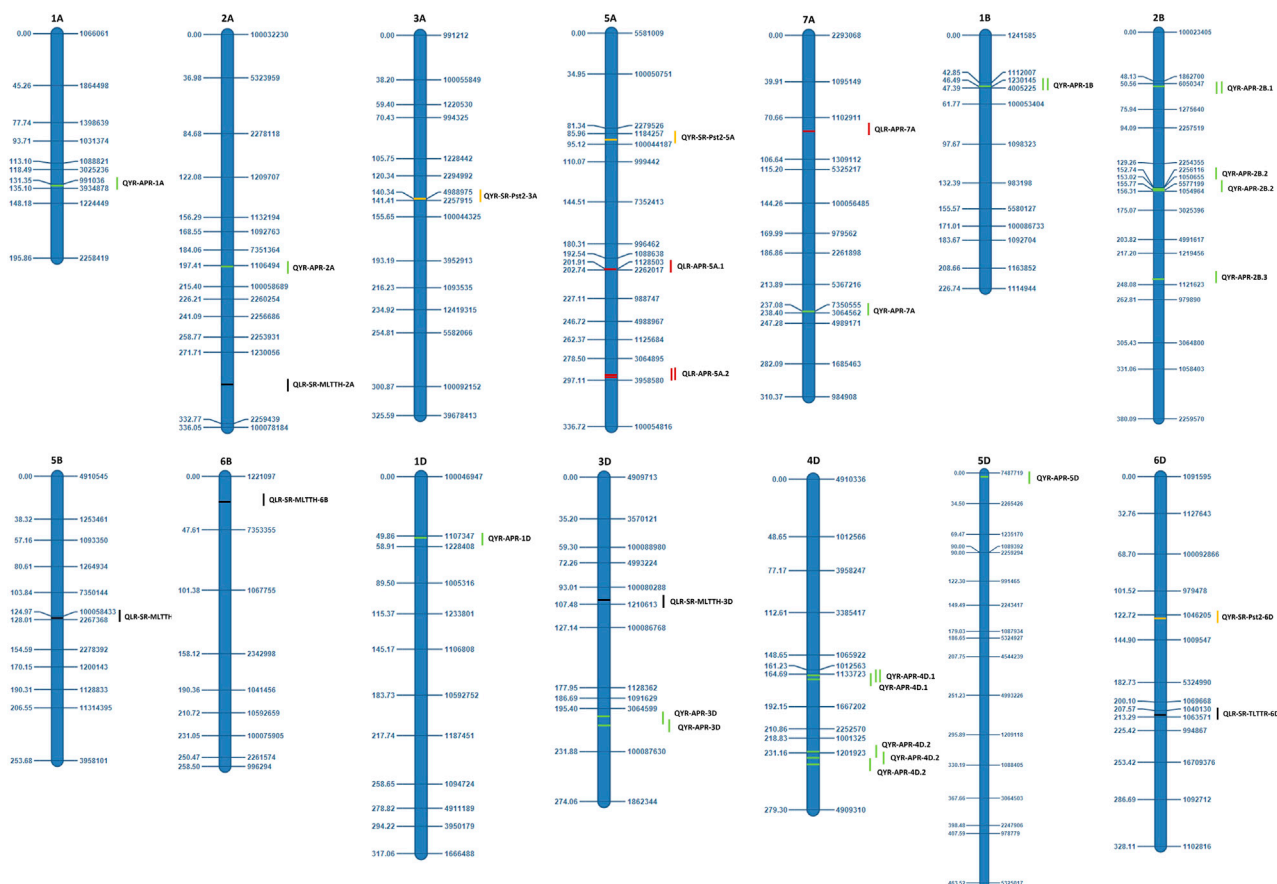


FIGURE 3

Identified QTLs and their genetic position from the Almay x Avocet population.

Stable QTLs for APR

A set of seven consistent QTLs including one for LR (*QLR-APR-5A.2*) and six for YR (*QYR-APR-1B*, *QYR-APR-2B.1*, *QYR-APR-2B.2*, *QYR-APR-3D*, *QYR-APR-4D.1*, and *QYR-APR-4D.2*) were identified in the Avocet population. The only consistent QTL, i.e., *QLR-APR-5A.2*, for leaf rust APR was identified in 2018 and across years, which was flanked between marker intervals 3064895–3958580 at a confidence interval of 284.5 cM–312.5 cM. The phenotypic variation explained was comparatively low, with 2.17% during 2018 and 1.74% across the years. A maximum of six consistent QTLs were identified for yellow rust APR. Two consistent QTLs, i.e., *QYR-APR-4D.1* and *QYR-APR-4D.2*, were identified in both the tested environments and also across years, which were flanked between the marker intervals 1133723–1667202 and 1201923–4909310 at confidence intervals 163.5 cM–176.5 cM and 224.5 cM–255.5 cM. Furthermore, these two QTLs reported high PVE which ranges from 5.58% to 8.55% (*QYR-APR-4D.1*) and 6.92% to 16.42% (*QYR-APR-4D.2*) (Table 5). The remaining four consistent QTLs were identified in one environment and also across years. Two QTLs, i.e., *QYR-APR-1B* and *QYR-APR-2B.2*, were identified in 2019 and across years, which were identified in the marker intervals of 1230145–4005225 (*QYR-APR-1B*) and 5577199–1054964 in 2018 and 2256116–1050655 across years (*QYR-APR-2B.2*) at a confidence interval of 46.5 cM–47.5 cM

and 152.5 cM–156.5 cM. The remaining two QTLs, i.e., *QYR-APR-2B.1* and *QYR-APR-3D*, were identified in 2018 and across years, which were mapped in the marker intervals of 48.5 cM–58.5 cM and 193.5 cM–218.5 cM, respectively. One QTL, i.e., *QYR-APR-3D*, also reported high PVE with 11.43% in 2018 and 12.24% across years.

QTLs for yellow and leaf rust ASR

A total of eight QTLs including five for LR and three for YR were identified in the Avocet population. For LR ASR, the highest PVE was reported for *QLR-ASR-TLTTT-6D* (7.58%), followed by *QLR-ASR-MLTTH-2A* (6.10%), *QLR-ASR-MLTTH-6B* (5.81%), *QLR-ASR-MLTTH-5B* (4.30%), and *QLR-ASR-MLTTH-3D* (0.82%). For yellow rust, the highest PVE was reported for *QYR-ASR-Pst2-5A* (14.01%), followed by *QYR-ASR-Pst2-6D* (5.80%) and *QYR-ASR-Pst2-3A* (4.75%).

Putative candidate genes

The putative candidate genes were identified for consistent QTLs with high PVE for leaf and yellow rust APR through *in silico* analysis and are presented in Table 6, and additional details are provided in Supplementary Table S4. The QTL, i.e., *QLR-APR-4A*, located at

TABLE 6 Putative candidate genes for leaf and yellow rust adult plant resistance.

Trait	QTL name	Population	Physical position (Mb)	TraesID	Putative candidate gene	Molecular function
LR_APR	QLR-APR-4A	Almaly x Anza	707.93–712.15	TraesCS4A02G443600.1	<i>Cytochrome P450</i>	Monooxygenase activity, iron ion binding, oxidoreductase activity, and heme binding
LR_APR	QLR-APR-3B	Almaly x Anza	22.24–444.82	TraesCS3B02G043400.1	<i>Leucine-rich repeat domain superfamily and protein kinase-like domain superfamily</i>	Protein kinase activity, nucleotide binding, protein binding, and transferase activity
YR_APR	QYR-APR-2A.1	Almaly x Anza	683.53–705.99	TraesCS2A02G458600.1	<i>Zinc-binding ribosomal protein</i>	Structural constituent of the ribosome and protein binding
YR_APR	QYR-APR-2A.2	Almaly x Anza	17.44–30.33	TraesCS2A02G069300.1	<i>SANT/Myb domain</i>	DNA binding
				TraesCS2A02G043700.1	<i>WRKY transcription factor</i>	DNA-binding transcription factor activity and sequence-specific DNA binding
YR_APR	QYR-APR-2B.1	Almaly x Avocet	NA-5.97	TraesCS2B02G011200.1	<i>Leucine-rich repeat-containing N-terminal and plant-type</i>	Protein binding
YR_APR	QYR-APR-2B.2	Almaly x Avocet	184.79–235.26	TraesCS2B02G222600.1	<i>Nucleotide-sugar transporter</i>	Pyrimidine nucleotide-sugar transmembrane transporter activity
YR_APR	QYR-APR-4D.1	Almaly x Avocet	481.88–495.07	TraesCS4D02G316800.1	<i>NAC domain superfamily</i>	DNA binding

LR_APR, leaf rust adult plant resistance; YR_APR, yellow rust adult plant resistance.

707.93–712.15 Mb encodes *cytochrome P450* (TraesCS4A02G443600.1). Similarly, QLR-APR-3B located at 22.24–444.82 Mb encodes the *leucine-rich repeat domain superfamily* (TraesCS3B02G043400.1). QYR-APR-2A.1 located at 683.53–705.99 Mb encodes *zinc-binding ribosomal protein* (TraesCS2A02G458600.1). QYR-APR-2A.2 located at 17.44–30.33 Mb encodes the *SANT/Myb domain* (TraesCS2A02G069300.1) and *WRKY transcription factor* (TraesCS2A02G043700.1). Other two QTLs, i.e., QYR-APR-4D.1 encodes the *NAC domain superfamily* (TraesCS4D02G316800.1) and QYR-APR-2B.2 located at 184.79–235.26 Mb encodes the *nucleotide-sugar transporter* (TraesCS2B02G222600.1).

Discussion

The present rate of genetic gain is approximately 0.8%–1.2% for the major food crops, including wheat, and in recent years this progress has plateaued. The current rate of annual progress is too short of the 2.4% required to feed approximately 9.5 billion people by 2050 (Krishnappa et al., 2021a; Ray et al., 2012; Ray et al., 2013). Although the genetic progress of crop plants is a continuous process, protection of crop yield from biotic and abiotic stresses is also very important to minimize the crop losses and to have sustainable crop production. Rust (yellow, leaf, and stem) diseases are very important biotic stresses in wheat, which cause substantial crop damage across the globe. Genetic dissection of complex traits through QTL mapping will be helpful in designing the appropriate breeding strategies through MAS (Krishnappa et al., 2021b; Khan et al., 2022). Many of the race-specific/seedling resistance genes identified for all three rusts are from wild relatives, and their direct utilization in breeding programs is hindered due to an undesirable linkage drag associated with resistance locus. Furthermore, the durability of the seedling resistance genes is less

compared to APR genes. Hence, to avoid linkage drag and resistance breakdown, plant breeders showed much interest in molecular studies in elite genetic backgrounds (Tehseen et al., 2022).

Phenotyping of 206 RILs from the Avocet and 160 RILs from Anza population suggests the presence of a wide variability of resistance to both the rusts, APR and ASR. Previously, a similar kind of broad variability was observed for wheat leaf rust (Rollar et al., 2021). A high broad sense heritability of approximately 0.85 and above was recorded for all the traits including leaf and yellow rust APR and leaf and yellow rust ASR in both populations. A similar range of broad sense heritability was also observed in previous reports (Rollar et al., 2021; Gao et al., 2016). The magnitude of correlations between leaf and yellow rust APR was relatively low, although a slight positive correlation was observed in the Anza population.

Overall, 51 QTLs in Anza (28 QTLs) and Avocet populations (23 QTLs) were identified. The 28 QTLs mapped in the Anza population represent leaf rust APR (6 QTLs), yellow rust APR (10 QTLs), leaf rust ASR (6 QTLs), and yellow rust ASR (6 QTLs). The 23 QTLs identified in the Avocet population represent leaf rust APR (3 QTLs), yellow rust APR (12 QTLs), leaf rust ASR (5 QTLs), and yellow rust ASR (3 QTLs). A maximum of 22 QTLs were mapped for yellow rust APR, followed by leaf rust ASR (11 QTLs), leaf rust APR, and ASR (6 QTLs each). Similarly, disease-wise representation of QTLs includes 31 QTLs for yellow rust and 20 QTLs for leaf rust. QTLs were equally distributed among the three subgenomes with 17 QTLs each. The maximum number of QTLs were mapped on chromosome 6B (6 QTLs) followed by 2A, 2B, 3A, 4D, and 6D (4 QTLs each); 3D, 5A, and 7A (3 QTLs each); 1A, 1B, 1D, 3B, 5B, and 7D (2 QTLs each); and 2D, 4A, 5D, and 7B (1 QTL each).

Nine QTLs were detected for leaf rust APR on eight chromosomes. Two major QTLs (QLR-APR-2B and QLR-APR-

7A) explained more than 11.0% PVE. Previously, one leaf rust APR gene, i.e., *Lr 34/Yr18*, was identified on the 7D chromosome on the Lalbahadur bread wheat cultivar through single-chromosome substitutions from the Parula cultivar, a source of *Lr34/Yr18* (Lagudah et al., 2006). Similarly, the 7D chromosome is important as many of the QTLs were detected previously on this chromosome (Gao et al., 2016; Li et al., 2016; Zhang et al., 2017; Bokore et al., 2020; Gebrewahid et al., 2020). We also detected two QTLs in 7D, namely, QLR-APR-7D and QYR-APR-7D, that had PVE of 4.81% and 7.34%, respectively. Similarly, several previous studies also reported QTLs for leaf rust APR on the same chromosomes 1A, 2B, 3B, 4A, 5A, 5B, and 7A at different locations (Gao et al., 2016; Li et al., 2016; Lan et al., 2017; Zhang et al., 2017; Zhang et al., 2019; Bokore et al., 2020; Gebrewahid et al., 2020; Rauf et al., 2022). For leaf rust seedling resistance, 11 QTLs were identified in both Anza and Avocet populations. Three major QTLs, i.e., QLR-ASR-TCPTQ-6D, QLR-ASR-TLTTR-6D, and QLR-ASR-THTTT-7A, had PVE of 14.63%, 7.58%, and 9.07%, respectively, and mapped at marker intervals 100,023455–1091595, 1040130–1063571, and 1111941–1125395 and confidence intervals 216.5 cM–228.5 cM, 209.5 cM–215.5 cM, and 137.5 cM–144.5 cM, respectively. Previously, several leaf rust ASR genes including *Lr2a*, *Lr3*, *Lr3a*, *Lr9*, *Lr11*, *Lr15*, *Lr17*, *Lr18*, *Lr20*, *Lr22a*, *Lr32*, *Lr36*, *Lr37*, *Lr38*, *Lr39*, *Lr41*, *Lr45*, *Lr47*, *Lr52*, *Lr53*, *Lr59*, 65, *Lr77*, *Lr79*, and *Lr80* were identified on chromosomes 2A, 2D, 3B, 3D, 5B, 6B, 6D, and 7A, respectively (Kumar et al., 2022); whereas in our study, 11 QTLs were mapped on the same chromosomes. Furthermore, previously, few leaf rust ASR QTLs were identified on the same chromosomes at different locations (Gao et al., 2016; Li et al., 2016; Zhang et al., 2019; Delfan et al., 2023). A maximum of 22 QTLs were found on different chromosomes covering all three subgenomes for yellow rust APR. A set of eight major yellow rust ASR QTLs had PVE $\geq 10.0\%$, with the highest variation explained by QYR-APR-3A.2 (20.93%) followed by QYR-APR-4D.2 (16.42%), QYR-APR-3D (12.24%), QYR-APR-3D (11.43%), QYR-APR-3A.3 (10.57%), QYR-APR-2A.2 (10.39%), QYR-APR-3A.1 (10.35%), and QYR-APR-1B (10.20%). Three yellow rust APR genes, i.e., *Yr49* and *Yr71* on 3D and *Yr75* on 7A chromosomes, were reported in previous studies (Jamil et al., 2020). Previously, a few yellow rust APR QTLs were also identified on the same chromosomes in different mapping populations and marker systems (Lan et al., 2017; Yuan et al., 2018; Long et al., 2019; Zhang et al., 2019; Gebrewahid et al., 2020; Farzand et al., 2021; Cheng et al., 2022; Tehseen et al., 2022). In our study, nine yellow rust ASR QTLs on chromosomes 1B, 3A, 5A, 6B, 6D, and 7B were identified. Out of nine QTLs, two major QTLs explained $\geq 8.0\%$ PVE, with the highest variation explained by QYR-ASR-*Pst1*-5A (14.01%) followed by QYR-ASR-*Pst1*-6D (8.57%).

Detection and validation of consistent QTLs in multiple environments are critical for their effective utilization through marker-based breeding approaches (Krishnappa et al., 2023). In this direction, Anza population was phenotyped for 3 consecutive years (2018, 2019, and 2020), whereas Avocet population was tested during 2 consecutive years (2018 and 2019). As a result, 13 consistent QTLs including nine QTLs for yellow rust APR (QYR-APR-2A.1, QYR-APR-2A.2, QYR-APR-4D.2, QYR-APR-1B, QYR-APR-2B.1, QYR-APR-2B.2, QYR-APR-3D, QYR-APR-4D.1, and QYR-APR-4D.2) and four QTLs for leaf rust APR (QLR-APR-

4A, QLR-APR-2B, QLR-APR-3B, and QLR-APR-5A.2) were identified. Stable QTLs are promising candidates for further validation in diverse backgrounds and exploitation in marker-assisted selection. Previously, Gebrewahid et al. (2020) reported QTLs on chromosomes 2BS and 5AL conferred resistance to both YR and LR and proposed that *QYr.hebau-5AL/QLr.hebau-5AL* are likely to be novel. Zhang et al. (2019) identified QTLs (*QLr.hebau-5AL/QYr.hebau-5AL*) on chromosome 5AL conferred resistance to both rusts; they are likely to be new QTLs. Bokore et al. (2023) reported that the wheat cultivar Carberry contributed QTLs conferring LR APR on 2B (two loci, i.e., *QLr.spa-2B.2* and *QLr.spa-2B.1*) and 5A (*QLr.spa-5A*). Kumar et al. (2020) identified three distinct loci revealed on chromosomes 2B (*QLr.ramp-2B.7* and *QLr.ramp-2B.8*) and 5A (*QLr.ramp-5A*) to be associated with LR-APR. QTLs associated with stripe rust resistance APR were identified on chromosome 3D (Ye et al., 2019; Habib et al., 2020). A significant association of IWA5707 and other linked SNPs (IWA6277, IWA5375, and IWA5766) was detected on the short arm of chromosome 4D at 25.7 cM. Muleta et al. (2017) and Forrest et al. (2014) reported a significant association of IWA5707, IWA5375, and IWA5766 on chromosome 4D with resistance to YR. A putatively new QTL, linked to LR APR, was identified on chromosome 4D (Rollar et al., 2021). Mapping results identified QTL-conferring APR to stripe rust resistance also on 4DL (Zhang et al., 2022). One QTL, i.e., *QLr.cimmyt-5A* associated with APR LR, was mapped on the long arm of chromosome 5A and closely linked to *Vrn-A1* at 587.0 Mb (Rosewarne et al., 2012).

The putative candidate genes for the QTLs are provided in Table 6; Supplementary Table S4. For instance, QLR-APR-4A encodes *cytochrome P450* (TraesCS4A02G443600.1) found to have a role in rust resistance, and transcription profiling suggests that transcripts encoding *cytochrome P450* were upregulated (Hulbert et al., 2007; Manickavelu et al., 2010; Wu et al., 2019) during rust pathogen infection. Similarly, QLR-APR-3B encodes the *leucine-rich repeat domain superfamily* (TraesCS3B02G043400.1) that is crucial for wheat rust resistance. The resistant hexaploid wheat variety Thatcher *Lr10* encodes a *nucleotide-binding site* (NBS) and *leucine-rich repeat* (LRR), which play a role in wheat leaf rust resistance (Feuillet et al., 2003). Most R-genes encode intracellular *nucleotide-binding leucine-rich-repeat receptors* (NBS-LRRs), which play a key role in wheat rust resistance (Hao et al., 2016; Basnet et al., 2022). Inactivation of the wheat Ser/Thr kinase gene, i.e., *Puccinia striiformis*-induced protein kinase 1 (TaPSPK1), results in broad-spectrum resistance to *Pst* races (Wang et al., 2022). The QTL, i.e., QYR-APR-2A.1, encodes *zinc-binding ribosomal protein* (TraesCS2A02G458600.1). Wheat zinc finger protein *TaLSD1* regulates a hypersensitive response in plants, thereby conferring stripe rust resistance (Guo et al., 2013). QYR-APR-2A.2 encodes the *SANT/Myb domain* (TraesCS2A02G069300.1). TFs including *Zn finger-binding proteins*, *SANT/Myb domains*, *NAC domain*, and *BTF3* play an important role in imparting stripe rust resistance (Jan et al., 2021). QYR-APR-2A.2 encodes the *WRKY transcription factor* (TraesCS2A02G043700.1). Wang et al. (2020) and Wang et al. (2016) reported the role of *TaWRKY70* in YR resistance, particularly during the seedling stage. Furthermore,

transgenic overexpression of barley WRKY genes, namely, *HvWRKY6* and *HvWRKY70*, confers YR resistance (Li et al., 2020). *QYR-APR-2B.2* encodes the *nucleotide-sugar transporter* (TraesCS2B02G222600.1), sugar transporters like *TaSTP6* (Huai et al., 2019), *TaSTP13* (Huai et al., 2020), and *PsHXT1* (Chang et al., 2020) are essential for the pathogenicity of the wheat rust pathogen, and it promotes wheat susceptibility to stripe rust. Another QTL *QYR-APR-4D.1* encodes the *NAC domain superfamily* (TraesCS4D02G316800.1), and wheat *NAC Transcription factors* like *TaNAC069* (Xu et al., 2022) and *TuNAC69* (Zhang et al., 2021) regulate leaf and stripe rust resistance, respectively. In crop plants, the majority of the disease-resistance genes is race-specific and contains the NBS and LRR domains. These resistant genes or QTLs are believed to be regulated by NBS domains through signal transduction, and the specific sites of corresponding pathogen virulence genes are recognized by LRR domains (Gill et al., 2015). Some of the stable QTLs like *QLR-APR-2B* (Anza population) and *QYR-APR-2B.2* (Avocet population) were found to encode *nucleotide-binding domains* which play a role in disease resistance. Similarly, other stable QTLs such as *QYR-APR-4D.2* (Anza population) and *QYR-APR-4D.1* (Avocet population) encode important putative genes like the *zinc finger C2H2 superfamily*, which play an important role in disease resistance in plants (Guo et al., 2013). Therefore, stable QTLs which encode the same putative candidate genes could be potential candidate genomic regions for further functional validation.

Conclusion

The study with two RIL populations derived from a cross between Almaly \times Anza (160 RILs) and Almaly \times Avocet S (206 RILs) suggested the presence of wide variability for yellow and leaf rust APR and ASR. We identified a set of 13 consistent QTLs including yellow rust APR (9 QTLs) and leaf rust APR (4 QTLs). Among them, *QLR-APR-2B* and *QYR-APR-4D.2* from the Anza population and *QLR-APR-5A.2*, *QYR-APR-4D.1*, *QYR-APR-4D.2*, and *QYR-APR-3D* from the Avocet population are important candidates to target for further validation and deployment in LR and YR resistance breeding. Several putative candidate genes were identified in this study; mainly, zinc finger proteins, DNA-binding pseudobarrel domain superfamily, and NAC domain superfamily with the associated functions in the resistance mechanism of leaf and yellow rust were identified. The functional characterization of these candidate genes will provide greater applicability of this study in rust resistance breeding.

Data availability statement

The phenotypic and genotypic datasets used in this study are available as Supplementary Material (Supplementary Table S1, Supplementary Table S2). Further, they were submitted to the open access repository “DRYAD” and they will be accessible through the link <https://doi.org/10.5061/dryad.3bk3j9krm>.

Author contributions

ALK: conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, resources, supervision, validation, visualization, writing–original draft, and writing–review and editing. NR: formal analysis, methodology, software, writing–original draft, and writing–review and editing. DS: conceptualization, data curation, formal analysis, methodology, software, and writing–review and editing. AM: data curation, investigation, software, and writing–original draft. MK: data curation, investigation, and writing–original draft. MN: conceptualization, investigation, software, and writing–original draft. AB: data curation, investigation, and writing–original draft. GK: data curation, software, and writing–review and editing. EG: data curation, investigation, methodology, and writing–original draft. AsK: investigation and writing–original draft. ZK: investigation and writing–original draft. KB: investigation and writing–original draft.

Funding

The author(s) declare that their financial support was received for the research, authorship, and/or publication of this article. This research has been funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (grant no. AP09258991 and grant no. BR18574099, the task 02).

Conflict of interest

NR was employed by Corteva Agriscience. DS was employed by Syngenta, Jealott’s Hill International Research Centre. The authors declare that no funding was obtained from either of the commercial organizations.

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/fgene.2023.1265859/full#supplementary-material>

References

- Aktar-Uz-Zaman, M., Tuhina-Khatun, M., Hanafi, M. M., and Sahebi, M. (2017). Genetic analysis of rust resistance genes in global wheat cultivars: an overview. *Biotechnol. Biotechnol. Equip.* 31 (3), 431–445. doi:10.1080/13102818.2017.1304180
- Ali, S., Gladieux, P., Leconte, M., Gautier, A., Justesen, A. F., Hovmoller, M. S., et al. (2014). Origin, migration routes and worldwide population genetic structure of the wheat yellow rust pathogen *Puccinia Striiformis* f. sp. *tritici*. *PLoS Pathog.* 10, e1003903. doi:10.1371/journal.ppat.1003903
- Basnet, B., Juliana, P., Bhattarai, K., and Upreti, U. (2022). A review on major rust resistance gene and amino acid changes on wheat (*Triticum aestivum* L.). *Adv. Agric.* 4, 1–11. doi:10.1155/2022/7419326
- Bokore, F. E., Knox, R. E., Cuthbert, R. D., Pozniak, C. J., McCallum, B. D., N'Diaye, A., et al. (2020). Mapping quantitative trait loci associated with leaf rust resistance in five spring wheat populations using single nucleotide polymorphism markers. *PLoS One* 15 (4), e0230855. doi:10.1371/journal.pone.0230855
- Chang, Q., Lin, X., Yao, M., Liu, P., Guo, J., Huang, L., et al. (2020). Hexose transporter PsHXT1-mediated sugar uptake is required for pathogenicity of wheat stripe rust. *Plant Biotechnol. J.* 18 (12), 2367–2369. doi:10.1111/pbi.13398
- Chen, X. (2013). High-temperature adult-plant resistance, key for sustainable control of stripe rust. *Am. J. Plant Sci.* 4, 608–627. doi:10.4236/ajps.2013.43080
- Cheng, B., Gao, X., Cao, N., Ding, Y., Chen, T., Zhou, Q., et al. (2022). QTL mapping for adult plant resistance to wheat stripe rust in M96-5 × Guixie 3 wheat population. *J. Appl. Genet.* 63, 265–279. doi:10.1007/s13353-022-00686-z
- Ciechanowska, I., Semagn, K., McCallum, B., Randhawa, H., Strenze, K., Dhariwal, R., et al. (2022). Quantitative trait locus mapping of rust resistance and agronomic traits in spring wheat. *Can. J. Plant Sci.* 102, 1139–1150. doi:10.1139/CJPS-2022-0023
- Delfan, S., Bihamta, M. R., Dadrezaei, S. T., Abbasi, A., and Alipour, H. (2023). Exploring genomic regions involved in bread wheat resistance to leaf rust at seedling/adult stages by using GWAS analysis. *BMC Genom.* 24, 83. doi:10.1186/s12864-022-09096-1
- Dreisigacker, S., Shewaryga, H., Crossa, J., Arief, V. N., DeLacy, I. H., Singh, R. P., et al. (2012). Genetic structures of the CIMMYT international yield trial targeted to irrigated environments. *Mol. Breed.* 29, 529–541. doi:10.1007/s11032-011-9569-7
- Edet, O. U., Gorafi, Y. S. A., Nasuda, S., and Tsujimoto, H. (2018). DArTseq-based analysis of genomic relationships among species of tribe Triticeae. *Sci. Rep.* 8, 16397. doi:10.1038/s41598-018-34811-y
- Ellis, J. G., Lagudah, E. S., Spielmeier, W., and Dodds, P. N. (2014). The past, present and future of breeding rust resistant wheat. *Front. Plant Sci.* 5, 641. doi:10.3389/fpls.2014.00641
- Farzand, M., Dhariwal, R., Hiebert, C. W., Spaner, D., and Randhawa, H. S. (2021). Mapping quantitative trait loci associated with stripe rust resistance from the Canadian wheat cultivar 'AAC Innova. *Can. J. Plant Pathol.* 43, S227–S241. doi:10.1080/07060661.2021.1982011
- Feuillet, C., Travella, S., Stein, N., Albar, L., Nublat, A., and Keller, B. (2003). Map-based isolation of the leaf rust disease resistance gene *Lr10* from the hexaploid wheat (*Triticum aestivum* L.) genome. *PNAS U. S. A.* 100 (25), 15253–15258. doi:10.1073/pnas.2435133100
- Flor, H. H. (1971). Current status of the gene-for-gene concept. *Annu. Rev. Phytopathol.* 9, 275–296. doi:10.1146/annurev.py.09.090171.001423
- Forrest, K., Pujol, V., Bulli, P., Pumphrey, M., Wellings, C., Herrera-Foessel, S., et al. (2014). Development of a SNP marker assay for the *Lr67* gene of wheat using a genotyping by sequencing approach. *Mol. Breed.* 34, 2109–2118. doi:10.1007/s11032-014-0166-4
- Gao, L., Turner, M. K., Chao, S., Kolmer, J., and Anderson, J. A. (2016). Genome wide association of seedling and adult plant resistance in elite spring wheat breeding lines. *PLoS One* 11, e0148671. doi:10.1371/journal.pone.0148671
- Gassner, G., and Straib, W. (1932). Untersuchungen über die Infektionsbedingungen von *Puccinia glumarum* und *Puccinia graminis*. *Arb. Biol. Reichsanst. Land-Forst-wirtsch. Berlin-Dahlem.* 16 (4), 609–629.
- Gebrewahid, T. W., Zhang, P., Zhou, Y., Yan, X., Xia, X., He, Z., et al. (2020). QTL mapping of adult plant resistance to stripe rust and leaf rust in a Fuyi 3/Zhengzhou 5389 wheat population. *Crop J.* 8 (4), 655–665. doi:10.1016/j.cj.2019.09.013
- Gill, U. S., Lee, S., and Mysore, K. S. (2015). Host versus non-host resistance: distinct wars with similar arsenals. *Phytopathology* 105 (5), 580–587. doi:10.1094/phyto-11-14-0298-rvw
- Godoy, J. G., Ryneerson, S., Chen, X., and Pumphrey, M. (2018). Genome-wide association mapping of loci for resistance to stripe rust in north American elite spring wheat germplasm. *Phytopathology* 108, 234–245. doi:10.1094/PHYTO-06-17-0195-R
- Guo, J., Bai, P., Yang, Q., Liu, F., Wang, X., Huang, L., et al. (2013). Wheat zinc finger protein TaLSD1, a negative regulator of programmed cell death, is involved in wheat resistance against stripe rust fungus. *Plant. Physiol. biochem.* 71, 164–172. doi:10.1016/j.plaphy.2013.07.009
- Habib, M., Awan, F. S., Sadia, B., and Zia, M. A. (2020). Genome-wide association mapping for stripe rust resistance in Pakistani spring wheat genotypes. *Plants* 9, 1056. doi:10.3390/plants9091056
- Hao, Y., Wang, T., Wang, K., Wang, X., Fu, Y., Huang, L., et al. (2016). Transcriptome analysis provides insights into the mechanisms underlying wheat plant resistance to stripe rust at the adult plant stage. *PLoS One* 11 (3), e0150717. doi:10.1371/journal.pone.0150717
- Herrera-Foessel, S. A., Singh, R. P., Huerta-Espino, J., Crossa, J., Yuen, J., and Djurle, A. (2006). Effect of leaf rust on grain yield and yield traits of durum wheats with race-specific and slow-rusting resistance to leaf rust. *Plant Dis.* doi:10.1094/PD-90-1065
- Huai, B., Yang, Q., Qian, Y., Qian, W., Kang, Z., and Liu, J. (2019). ABA-Induced sugar transporter TaSTP6 promotes wheat susceptibility to stripe rust. *Plant Physiol.* 181 (3), 1328–1343. doi:10.1104/pp.19.00632
- Huai, B., Yang, Q., Wei, X., Pan, Q., Kang, Z., and Liu, J. (2020). TaSTP13 contributes to wheat susceptibility to stripe rust possibly by increasing cytoplasmic hexose concentration. *BMC Plant Biol.* 20 (1), 49. doi:10.1186/s12870-020-2248-2
- Huerta-Espino, J., Singh, R. P., Germán, S., McCallum, B. D., Park, R. F., Chen, W. Q., et al. (2011). Global status of wheat leaf rust caused by *Puccinia trititica*. *Euphytica* 179, 143–160. doi:10.1007/s10681-011-0361-x
- Hulbert, S. H., Bai, J., Fellers, J. P., Pacheco, M. G., and Bowden, R. L. (2007). Gene expression patterns in near isogenic lines for wheat rust resistance gene *Lr34/Yr18*. *Phytopathology* 97 (9), 1083–1093. doi:10.1094/PHYTO-97-9-1083
- International Wheat Genome Sequencing Consortium (IWGSC) (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* 361 (6403). doi:10.1126/science.aar7191
- Jamil, S., Shahzad, R., Ahmad, S., Fatima, R., Zahid, R., Anwar, M., et al. (2020). Role of genetics, genomics, and breeding approaches to combat stripe rust of wheat. *Front. Nutr.* 7, 580715. doi:10.3389/fnut.2020.580715
- Jan, I., Saripalli, G., Kumar, K., Kumar, A., Singh, R., Batra, R., et al. (2021). Meta-QTLs and candidate genes for stripe rust resistance in wheat. *Sci. Rep.* 11 (1), 22923. doi:10.1038/s41598-021-02049-w
- Johnson, R., Stubbs, R. W., Fuchs, E., and Chamberlain, N. H. (1972). Nomenclature for physiological races of *Puccinia striiformis* infecting wheat. *Trans. Br. Mycol. Soc.* 58, 475–480. doi:10.1016/S0007-1536(72)80096-2
- Jones, J. D. G., and Dangl, J. L. (2006). The plant immune system. *Nature* 444, 323–329. doi:10.1038/nature05286
- Khan, H., Krishnappa, G., Kumar, S., Mishra, C. N., Krishna, H., Devate, N. B., et al. (2022). Genome-wide association study for grain yield and component traits in bread wheat (*Triticum aestivum* L.). *Front. Genet.* 13, 982589. doi:10.3389/fgene.2022.982589
- Koishybaev, M. K., Zhanarbekova, A. B., Kokhmetova, A. M., and Rsaiev, Sh. S. (2010). Genetic study of wheat resistance to leaf rust. *Izvestiya NAS RK. Ser. Biol. Med.* 6, 10–15.
- Kokhmetova, A., Madenova, M., Purnhauser, L., Kampitova, G., Urzaliev, R., and Yessimbekova, M. (2016). Identification of leaf rust resistance genes in wheat cultivars produced in Kazakhstan. *Cereal Res. Commun.* 44, 240–250. doi:10.1556/0806.43.2015.056
- Kokhmetova, A., Rsaiev, A., Malysheva, A., Atishova, M., Kumarbayeva, M., and Keishilov, Z. (2021a). Identification of stripe rust resistance genes in common wheat cultivars and breeding lines from Kazakhstan. *Plants* 10, 2303. doi:10.3390/plants10112303
- Kokhmetova, A., Rsaiev, S., Atishova, M., Kumarbayeva, M., Malysheva, A., Keishilov, Z., et al. (2021b). Evaluation of wheat germplasm for resistance to leaf rust (*Puccinia trititica*) and identification of the sources of Lr resistance genes using molecular markers. *Plants* 10 (7), 1484. doi:10.3390/plants10071484
- Kokhmetova, A., Sharma, R., Rsaiev, S., Galymbek, K., Baymagambetova, K., Ziyaev, Z., et al. (2018). Evaluation of Central Asian wheat germplasm for stripe rust resistance. *Plant Genet. Resour.* 16, 178–184. doi:10.1017/S1479262117000132
- Kolmer, J. A., Bajgain, P., Rouse, M. N., Li, J., and Zhang, P. (2023). Mapping and characterization of the recessive leaf rust resistance gene *Lr83* on wheat chromosome arm 1DS. *Theor. Appl. Genet.* 136 (5), 115. doi:10.1007/s00122-023-04361-7
- Kolmer, J. A., Kابدulova, M. G., Mustafina, M. A., Zhemchuzhina, N. S., and Dubovoy, V. (2014). Russian populations of *Puccinia trititica* in distant regions are not differentiated for virulence and molecular genotype. *Plant Pathol.* 64, 328–336. doi:10.1111/ppa.12248
- Kolmer, J. A., and Ordóñez, M. E. (2007). Genetic differentiation of *Puccinia trititica* populations in central Asia and the caucasus. *Phytopathology* 97, 1141–1149. doi:10.1094/PHYTO-97-9-1141
- Kolmer, J. A. A. (2015). QTL on chromosome 5BL in wheat enhances leaf rust resistance of Lr46. *Mol. Breed.* 35, 74. doi:10.1007/s11032-015-0274-9
- Kosambi, D. D. (1943). The estimation of map distance from recombination values. *Ann. Eugen.* 12, 172–175. doi:10.1111/j.1469-1809.1943.tb02321.x

- Krattinger, S. G., Sucher, J., Selter, L. L., Chauhan, H., Zhou, B., Tang, M. Z., et al. (2016). The wheat durable, multipathogen resistance gene *Lr34* confers partial blast resistance in rice. *Plant Biotechnol.* 14, 1261–1268. doi:10.1111/pbi.12491
- Krishnappa, G., Khan, H., Krishna, H., Devate, N. B., Kumar, S., Mishra, C. N., et al. (2023). Genome-wide association study for grain protein, thousand kernel weight, and normalized difference vegetation index in bread wheat (*Triticum aestivum* L.). *Genes* 14, 637. doi:10.3390/genes14030637
- Krishnappa, G., Rathana, N. D., Sehgal, D., Ahlawat, A. K., Singh, S. K., Singh, S. K., et al. (2021b). Identification of novel genomic regions for biofortification traits using an SNP marker-enriched linkage map in wheat (*Triticum aestivum* L.). *Front. Nutr.* 8, 669444. doi:10.3389/fnut.2021.669444
- Krishnappa, G., Savadi, S., Tyagi, B. S., Singh, S. K., Mamrutha, H. M., Kumar, S., et al. (2021a). Integrated genomic selection for rapid improvement of crops. *Genomics* 113, 1070–1086. doi:10.1016/j.ygeno.2021.02.007
- Kthiri, D., Loladze, A., N'Diaye, A., Nilsen, K. T., Walkowiak, S., Dreisigacker, S., et al. (2019). Mapping of genetic loci conferring resistance to leaf rust from three globally resistant durum wheat sources. *Front. Plant Sci.* 10, 1247. doi:10.3389/fpls.2019.01247
- 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, 748. doi:10.3389/fpls.2020.00748
- Kumar, K., Jan, I., Saripalli, G., Sharma, P. K., Mir, R. R., Balyan, H. S., et al. (2022). An update on resistance genes and their use in the development of leaf rust resistant cultivars in wheat. *Front. Genet.* 13, 816057. doi:10.3389/fgene.2022.816057
- Lagudah, E. S., McFadden, H., Singh, R. P., Huerta-Espino, J., Bariana, H. S., and Spielmeier, W. (2006). Molecular genetic characterization of the *Lr34/yr18* slow rusting resistance gene region in wheat. *Theor. Appl. Genet.* 114 (1), 21–30. doi:10.1007/s00122-006-0406-z
- Lan, C., Hale, I. L., Herrera-Foessel, S. A., Basnet, B. R., Randhawa, M. S., Huerta-Espino, J., et al. (2017). Characterization and mapping of leaf rust and stripe rust resistance loci in hexaploid wheat lines UC1110 and PI610750 under Mexican environments. *Front. Plant Sci.* 8, 1450. doi:10.3389/fpls.2017.01450
- Li, C., Bai, G., Carver, B. F., Chao, S., and Wang, Z. H. (2016). Mapping quantitative trait loci for plant adaptation and morphology traits in wheat using single nucleotide polymorphisms. *Euphytica* 208, 299–312. doi:10.1007/s10681-015-1594-x
- Li, H., Li, Z., and Wang, J. (2008). Inclusive composite interval mapping (ICIM) for digenic epistasis of quantitative traits in biparental populations. *Theor. Appl. Genet.* 116, 243–260. doi:10.1007/s00122-007-0663-5
- Li, H., Wu, J., Shang, X., Geng, M., Gao, J., Zhao, S., et al. (2020). WRKY transcription factors shared by BTH-induced resistance and NPR1-mediated acquired resistance improve broad-spectrum disease resistance in wheat. *Mol. Plant. Microbe. Interact.* 33 (3), 433–443. doi:10.1094/MPMI-09-19-0257-R
- Liu, L., Wang, M. N., Feng, J. Y., See, D. R., Chao, S. M., and Chen, X. M. (2018). Combination of all-stage and high-temperature adult-plant resistance QTL confers high-level, durable resistance to stripe rust in winter wheat cultivar Madsen. *Theor. Appl. Genet.* 131 (9), 1835–1849. doi:10.1007/s00122-018-3116-4
- Long, D. L., and Kolmer, J. A. A. (1989). North American system of nomenclature for *Puccinia recondita* f. sp. *tritici*. *Phytopathology* 79, 525–529.
- Long, L., Yao, F., Yu, C., Ye, X., Cheng, Y., Wang, Y., et al. (2019). Genome-wide association study for adult-plant resistance to stripe rust in Chinese wheat landraces (*Triticum aestivum* L.) from the yellow and huai river valleys. *Front. Plant Sci.* 10, 596. doi:10.3389/fpls.2019.00596
- Mains, E. B., and Jackson, H. S. (1926). Physiologic specialization in the leaf rust of wheat, *Puccinia triticina* Erikss. *Phytopathology* 16, 89–120.
- Manickavelu, A., Kawaura, K., Oishi, K., Shin-I, T., Kohara, Y., Yahiaoui, N., et al. (2010). Comparative gene expression analysis of susceptible and resistant near-isogenic lines in common wheat infected by *Puccinia triticina*. *DNA Res.* 4, 211–222. doi:10.1093/dnares/dsq009
- McIntosh, R., Dubcovsky, J., Rogers, W., Xia, X., and Raupp, W. (2020). Catalogue of gene symbols for wheat: 2020 supplement. *Ann. Wheat Newsl.* 66, 109–128.
- McIntosh, R. A., Dubcovsky, J., Rogers, W. J., Morris, C., Appels, R., and Xia, X. C. (2016). Catalogue of gene symbols for wheat: 2015–2016 supplement. Available at: <https://shigen.nig.ac.jp/wheat/komugi/genes/macgene/supplement2015.pdf>.
- McIntosh, R. A., Wellings, C. R., and Park, R. F. (1995). *Wheat rusts: an atlas of resistance genes*. Australia: CSIRO.
- Morgounov, A., Tufan, H. A., Sharma, R., Akin, B., Bagci, A., Braun, H. J., et al. (2013). 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
- Muleta, K. T., Bulli, P., Ryneerson, S., Chen, X., and Pumphrey, M. (2017). Loci associated with resistance to stripe rust (*Puccinia striiformis* f. sp. *tritici*) in a core collection of spring wheat (*Triticum aestivum*). *PLoS One* 12, e0179087. doi:10.1371/journal.pone.0179087
- Peterson, R. F., Campbell, A. B., and Hannah, A. E. A. (1948). Diagrammatic scale for estimating rust intensity on leaves and stems of cereals. *Canad. J. Res.* 26, 496–500. doi:10.1139/cjr48c-033
- Randhawa, H., Puchalski, B. J., Frick, M., Goyal, A., Despins, T., Graf, R. J., et al. (2012). Stripe rust resistance among western Canadian spring wheat and triticales varieties. *Can. J. Plant. Sci.* 92 (4), 713–722. doi:10.4141/cjps2011-252
- Rathan, N. D., Krishnappa, G., Singh, A. M., and Govindan, V. (2023). Mapping QTL for phenological and grain-related traits in a mapping population derived from high-zinc-biofortified wheat. *Plants* 12 (1), 220. doi:10.3390/plants12010220
- Rauf, Y., Lan, C., Randhawa, M., Singh, R. P., Huerta-Espino, J., and Anderson, J. A. (2022). Quantitative trait loci mapping reveals the complexity of adult plant resistance to leaf rust in spring wheat 'Copio. *Crop Sci.* 62 (3), 1037–1050. doi:10.1002/csc.20728
- 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 (6), e66428. doi:10.1371/journal.pone.0066428
- Ray, D. K., Ramankutty, N., Mueller, N. D., and WestFoley, P. C. J. A. (2012). Recent patterns of crop yield growth and stagnation. *Nat. Commun.* 3, 1293. doi:10.1038/ncomms2296
- Roelfs, A. P., Singh, R. P., and Saari, E. E. (1992). *Rust diseases of wheat: Concepts and methods of disease management*. Mexico, D.F.: CIMMYT, 81.
- Rollar, S., Serfling, A., Geyer, M., Hartl, L., Mohler, V., and Ordon, F. (2021). QTL mapping of adult plant and seedling resistance to leaf rust (*Puccinia triticina* Erikss.) in a multiparent advanced generation intercross (MAGIC) wheat population. *Theor. Appl. Genet.* 134 (1), 37–51. doi:10.1007/s00122-020-03657-2
- Rosewarne, G. M., Singh, R. P., Huerta-Espino, J., Herrera-Foessel, S. A., Forrest, K. L., Hayden, M. J., et al. (2012). Analysis of leaf and stripe rust severities reveals pathotype changes and multiple minor QTLs associated with resistance in an Avocet × Pastor wheat population. *Theor. Appl. Genet.* 124 (7), 1283–1294. doi:10.1007/s00122-012-1786-x
- Sansaloni, C., Petroli, C., Jaccoud, D., Carling, J., and Detering, F. (2011). Diversity arrays technology (DARt) and next-generation sequencing combined: genomewide, high throughput, highly informative genotyping for molecular breeding of Eucalyptus. *BMC Proc.* 5 (54). doi:10.1186/1753-6561-5-S7-P54
- Schachtel, G. A., Dinooor, A., Herrmann, A., and Kosman, E. (2012). Comprehensive evaluation of virulence and resistance data: a new analysis Tool. *Plant Dis.* 96, 1060–1063. doi:10.1094/PDIS-02-12-0114-SR
- Sharma, R. C., Morgounov, A., Akin, B., Bespalova, L., Lang, L., Litvinenko, M., et al. (2014). Winter wheat East European regional yield trial: identification of superior genotypes and characterization of environments. *Crop Sci.* 54, 2469–2480. doi:10.2135/cropsci2014.01.0028
- Sharma-Poudyal, D., Chen, X. M., Wan, A. M., Zhan, G. M., Kang, Z. S., Cao, S. Q., et al. (2013). Virulence characterization of international collections of the wheat stripe rust pathogen, *Puccinia striiformis* f. sp. *tritici*. *Plant Dis.* 97, 379–386. doi:10.1007/s00122-023-04374-2doi:10.1094/PDIS-01-12-0078-RE
- Singh, R. P., Huerta-Espino, J., and Rajaram, S. (2000). Achieving near immunity to leaf and stripe rusts in wheat by combining slow rusting resistance genes. *Acta Phytopathologica Entomologica Hung.* 35, 133–139.
- Singh, S., and Bowden, R. L. (2011). Molecular mapping of adult-plant race-specific leaf rust resistance gene *Lr12* in bread wheat. *Mol. Breed.* 28, 137–142. doi:10.1007/s11032-010-9467-4
- Tehseen, M. M., Tonk, F. A., Tosun, M., Randhawa, H. S., Kurtulus, E., Ozseven, I., et al. (2022). QTL mapping of adult plant resistance to stripe rust in a doubled haploid wheat population. *Front. Genet.* 13, 900558. doi:10.3389/fgene.2022.900558
- Wang, H., Zou, S., Li, Y., Lin, F., and Tang, D. (2020). An ankyrin-repeat and WRKY-domain-containing immune receptor confers stripe rust resistance in wheat. *Nat. Commun.* 11, 1353. doi:10.1038/s41467-020-15139-6
- Wang, J., Li, Z., Shi, L., Zhu, L., Ren, Z., Li, X., et al. (2015). QTL mapping for adult-plant leaf rust resistance genes in Chinese wheat cultivar weimai 8. *Czech J. Genet. Plant Breed.* 51 (3), 79–85. doi:10.17221/51/2015-CJGPB
- Wang, J., Tao, F., An, F., Zou, Y., Tian, W., Chen, X., et al. (2016). Wheat transcription factor TaWRKY70 is positively involved in high-temperature seedling plant resistance to *Puccinia striiformis* f. sp. *tritici*. *Mol. Plant Pathol.* doi:10.1111/mpp.12425
- Wang, J. K., Li, H. H., Zhang, L. Y., and Meng, L. (2012). "QTL IciMapping version 3.2," in *Quantitative genetics group, inst. Of crop sci., Chinese acad. Of Agric.Sci.* (Beijing. Available at: <http://www.isbreeding.net> (Accessed September 16, 2013).
- Wang, N., Tang, C., Fan, X., He, M., Gan, P., Zhang, S., et al. (2022). Inactivation of a wheat protein kinase gene confers broad-spectrum resistance to rust fungi. *Cell* 185 (16), 2961–2974. doi:10.1016/j.cell.2022.06.027
- Wellings, C. R. (2011). Global status of stripe rust: a review of historical and current threats. 179(1), 129–141. doi:10.1007/s10681-011-0360-y
- Wu, J., Gao, J., Bi, W., Zhao, J., Yu, X., Li, Z., et al. (2019). Genome-wide expression profiling of genes associated with the *Lr47*-mediated wheat resistance to leaf rust (*Puccinia triticina*). *Int. J. Mol. Sci.* 20 (18), 4498. doi:10.3390/ijms20184498
- Xu, Y., Zou, S., Zeng, H., Wang, W., Wang, B., Wang, H., et al. (2022). NAC transcription factor TuNAC69 contributes to ANK-NLR-WRKY NLR-mediated stripe rust resistance in the diploid wheat *Triticum urartu*. *Int. J. Mol. Sci.* 23 (1), 564. doi:10.3390/ijms23010564
- Ye, X., Li, J., Cheng, Y., Yao, F., Long, L., Yu, C., et al. (2019). Genome-wide association study of resistance to stripe rust (*Puccinia striiformis* f. sp. *tritici*) in Sichuan wheat. *BMC Plant Biol.* 19 (1), 147. doi:10.1186/s12870-019-1764-4

- Yessenbekova, G. T., Kokhmetova, A. M., Kampitova, G. A., and Radivoje, J. (2016). Sources and donors for soft wheat selection by resistance to yellow rust. *Biosci. Biotechnol. Res. Asia* 13 (2), 693–700. doi:10.13005/bbra/2086
- Yuan, F.-P., Zeng, Q.-D., Wu, J.-H., Wang, Q.-L., Yang, Z.-J., Liang, B.-P., et al. (2018). QTL mapping and validation of adult plant resistance to stripe rust in Chinese wheat landrace humai 15. *Front. Plant Sci.* 9, 968. doi:10.3389/fpls.2018.00968
- Zhang, C., Huang, L., Zhang, H., Hao, Q., Lyu, B., Wang, M., et al. (2019). An ancestral NB-LRR with duplicated 3' UTRs confers stripe rust resistance in wheat and barley. *Nat. Commun.* 10 (1), 1–12. doi:10.1038/s41467-019-11872-9
- Zhang, P., Lan, C., Singh, R. P., Huerta-Espino, J., Li, Z., Lagudah, E., et al. (2022). Identification and characterization of resistance loci to wheat leaf rust and stripe rust in Afghan landrace “KU3067”. *Front. Plant Sci.* 13, 894528. doi:10.3389/fpls.2022.894528
- Zhang, P., Li, X., Gebrewahid, T. W., Liu, H., Xia, X., He, Z., et al. (2019). QTL mapping of adult-plant resistance to leaf and stripe rust in wheat cross SW 8588/thatcher using the wheat 55K SNP array. *Plant Dis.* 103 (12), 3041–3049. doi:10.1094/PDIS-02-19-0380-RE
- Zhang, P., Yin, G., Zhou, Y., Qi, A., Gao, F., Xia, X., et al. (2017). QTL mapping of adult-plant resistance to leaf rust in the wheat cross zhou 8425B/Chinese spring using high-density SNP markers. *Front. Plant Sci.* 8, 793. doi:10.3389/fpls.2017.00793
- Zhang, Y., Geng, H., Cui, Z., Wang, H., and Liu, D. (2021). Functional analysis of wheat NAC transcription factor, TaNAC069, in regulating resistance of wheat to leaf rust fungus. *Front. Plant Sci.* 12, 604797. doi:10.3389/fpls.2021.604797
- Zhu, Z., Cao, Q., Han, D., Wu, J., Wu, L., Tong, J., et al. (2023). Molecular characterization and validation of adult-plant stripe rust resistance gene Yr86 in Chinese wheat cultivar Zhongmai 895. *Theor. Appl. Genet.* 136 (6), 142.
- Ziyayev, Z. M., Sharma, R. C., Nazari, K., Morgounov, A. I., Amanov, A. A., Ziyadullaev, Z. F., et al. (2011). Improving wheat stripe rust resistance in Central Asia and the Caucasus. *Euphytica* 179, 197–207. doi:10.1007/s10681-010-0305-x



OPEN ACCESS

EDITED BY

Amaranatha Reddy Vennapusa,
Delaware State University, United States

REVIEWED BY

Chunjie Fan,
Chinese Academy of Forestry, China
Vishnutej Ellur,
Syngenta, United States

*CORRESPONDENCE

Songbai Zhang,
✉ yangtze2008@126.com
Ying Li,
✉ liying03@caas.cn

[†]These authors have contributed equally to
this work

RECEIVED 07 September 2023

ACCEPTED 21 December 2023

PUBLISHED 08 January 2024

CITATION

Song H, Gao X, Song L, Jiao Y, Shen L, Yang J,
Li C, Shang J, Wang H, Zhang S and Li Y (2024),
Unraveling the regulatory network of miRNA
expression in Potato Y virus-infected of
Nicotiana benthamiana using integrated small
RNA and transcriptome sequencing.
Front. Genet. 14:1290466.
doi: 10.3389/fgene.2023.1290466

COPYRIGHT

© 2024 Song, Gao, Song, Jiao, Shen, Yang, Li,
Shang, Wang, Zhang and Li. This is an open-
access article distributed under the terms of the
[Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/).
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.

Unraveling the regulatory network of miRNA expression in Potato Y virus-infected of *Nicotiana benthamiana* using integrated small RNA and transcriptome sequencing

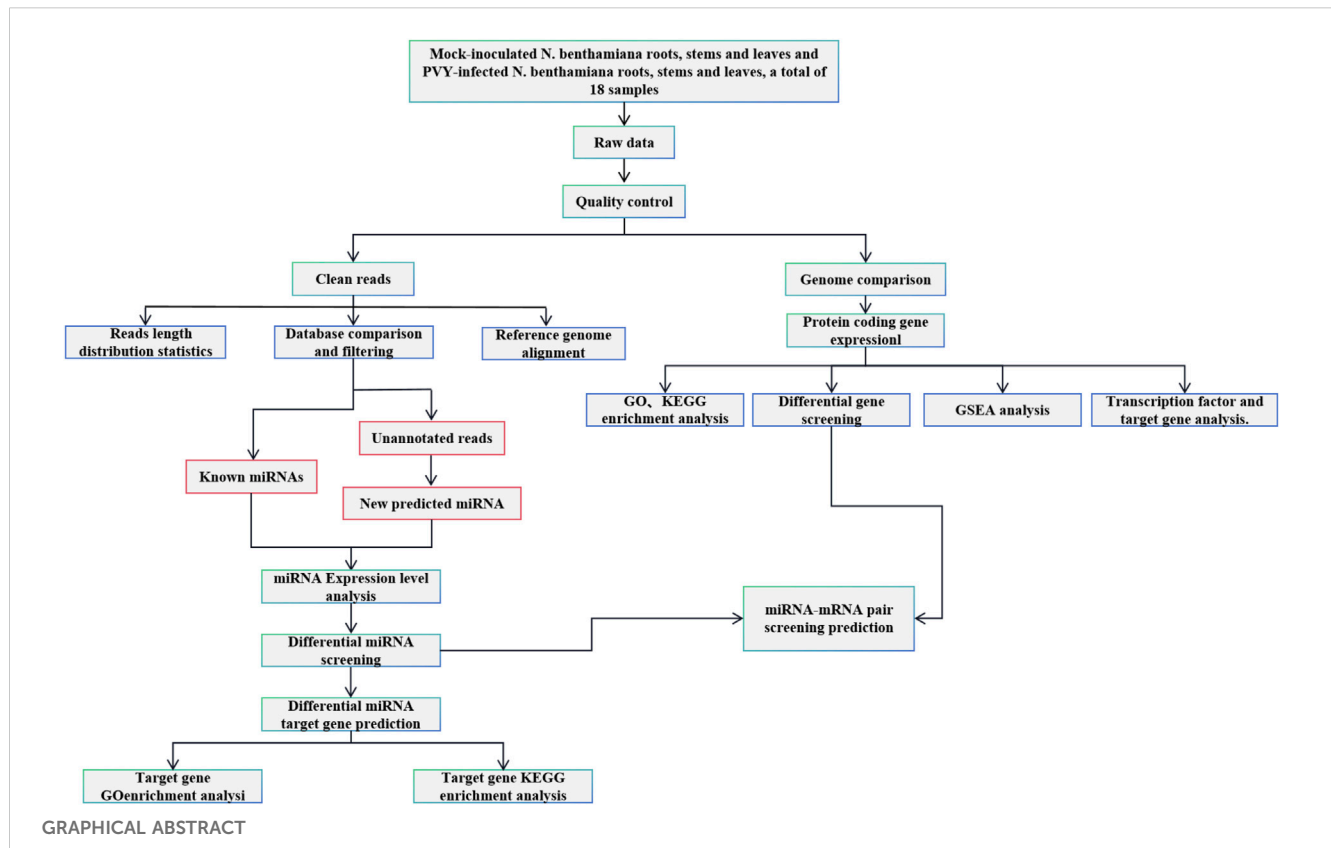
Hongping Song^{1†}, Xinwen Gao^{2†}, Liyun Song², Yubing Jiao²,
Lili Shen², Jinguang Yang², Changquan Li³, Jun Shang³,
Hui Wang⁴, Songbai Zhang^{1*} and Ying Li^{2*}

¹Hubei Engineering Research Center for Pest Forewarning and Management, Yangtze University, Jingzhou, Hubei, China, ²Key Laboratory of Tobacco Pest Monitoring Controlling and Integrated Management, Tobacco Research Institute of Chinese Academy of Agricultural Sciences, Qingdao, China, ³Liupanshui City Company of Guizhou Tobacco Company, Guizhou, China, ⁴Luoyang City Company of Henan Tobacco Company, Luoyang, Henan, China

Potato virus Y (PVY) disease is a global problem that causes significant damage to crop quality and yield. As traditional chemical control methods are ineffective against PVY, it is crucial to explore new control strategies. MicroRNAs (miRNAs) play a crucial role in plant and animal defense responses to biotic and abiotic stresses. These endogenous miRNAs act as a link between antiviral gene pathways and host immunity. Several miRNAs target plant immune genes and are involved in the virus infection process. In this study, we conducted small RNA sequencing and transcriptome sequencing on healthy and PVY-infected *N. benthamiana* tissues (roots, stems, and leaves). Through bioinformatics analysis, we predicted potential targets of differentially expressed miRNAs using the *N. benthamiana* reference genome and the PVY genome. We then compared the identified differentially expressed mRNAs with the predicted target genes to uncover the complex relationships between miRNAs and their targets. This study successfully constructed a miRNA-mRNA network through the joint analysis of Small RNA sequencing and transcriptome sequencing, which unveiled potential miRNA targets and identified potential binding sites of miRNAs on the PVY genome. This miRNA-mRNA regulatory network suggests the involvement of miRNAs in the virus infection process.

KEYWORDS

high-throughput sequencing, miRNA, Potato virus Y, miRNA-mRNA network, viral infection



1 Introduction

Plant viruses are intracellular parasites and pathogens that consist of nucleic acid, either single- or double-stranded RNA or DNA. They rely entirely on the host cell for replication and dissemination. When a plant is infected by a virus, it can cause disease-like physiological disorders that may lead to plant death. These viral diseases have a significant negative impact on agricultural production, resulting in crop yield losses and posing a threat to global agricultural production (Boualem et al., 2016). Among these pathogens, Potato virus Y (PVY) is particularly detrimental as it causes substantial reductions in both crop quantity and quality (Gray et al., 2010). Persistent PVY infection can lead to a decline in yield ranging from 40% to 60% (Whitworth et al., 2006). Plant RNA viruses, due to their evolvability, large population size, error-prone replication, and strong dependence on host organisms (Qin et al., 2023), have a significant impact on agricultural systems. To date, enhancing crop resistance is the most effective strategy for combating viral infections (Meziadi et al., 2017). Upon viral infection, plants activate a sophisticated antiviral immune response to combat the invader. At the same time, when viruses infect plants, the changes in host genes in different tissues are also different (Naqvi et al., 2011). This may be caused by the different functions of different tissues during virus infection.

MiRNA is the second most abundant sRNA in plants. miRNAs are non-coding RNAs ranging in length from 21 to 24 nucleotides (nt). They function by binding to target mRNA, resulting in translational inhibition or RNA destabilization. As a result, they

regulate gene expression at the post-transcriptional level (Davis et al., 2006). Most eukaryotic MIR genes are transcribed by RNA polymerase II (Pol II) to produce primary miRNA transcripts known as pri-miRNAs. These pri-miRNAs have imperfectly folded structures and undergo processing to form stem-loop precursors (pre-miRNAs). The pre-miRNAs are then cleaved to form RNA duplexes. The activation of RNAi occurs through specific double-stranded RNA, which is then incorporated into Argonaute proteins and assembled into RNA-induced silencing complexes (RISCs). These RISCs target mRNA or DNA that is complementary to small RNA (sRNA), thereby exerting regulatory control by binding to and cleaving mRNA, inhibiting translation initiation, and inducing DNA methylation (Ghildiyal and Zomare, 2009; Dang et al., 2011). This intricate process generates various sRNAs, including microRNAs (miRNAs) and small interfering RNAs (siRNAs) (Bologna and Voinnet, 2014). Additionally, miRNAs play significant roles in diverse biological processes, including plant growth, hormone secretion, signaling, protein degradation, and defense against biotic and abiotic stresses (Voinnet, 2009).

Meanwhile, miRNAs play a crucial role in plant-virus interactions (Mengistu and Tenkegna, 2021). To counteract host defense mechanisms, viruses employ various strategies to express viral suppressor genes that inhibit RNA silencing (RSV) during infection (Li et al., 2012; Zhao and Guo, 2022). Conversely, plants activate specific defenses upon viral infection to counteract RNA silencing inhibition (Ding and Voinnet, 2007). Plant miRNAs have the ability to suppress viral expression, induce cleavage of viral mRNA, or impede its translation (Liu SR. et al., 2017; Liu WW. et al.,

2017). For example, the endogenous miR393 in plants was the first identified miRNA that enhanced disease resistance by inhibiting the auxin signaling pathway (Navarro et al., 2006). In tobacco plants, miR482 was found to regulate the R gene N of Tobacco mosaic virus (TMV) thereby modulating viral infection by targeting disease-resistant proteins (Shivaprasad et al., 2012; Yang et al., 2015), and overexpression of miR6019 resulted in reduced N transcript accumulation and impaired N-mediated TMV resistance (Li et al., 2012). In previous studies, researchers utilized miR159, miR167b, and miR171 precursors as the basis for designing targeting sequences using artificial miRNA (amiRNA) technology, conferred resistance against PVY and Potato virus X in tobacco plants (Ai et al., 2011). These findings emphasize the critical regulatory role of miRNAs in the viral infection process, exerting direct or indirect control over virus infection dynamics.

Recent investigations have utilized antiviral miRNAs to design and generate amiRNAs, which effectively silence target genes or viruses in plants and achieve robust virus resistance (Duan et al., 2008; Niu et al., 2006; Fahim et al., 2012; Zhang et al., 2018). Therefore, miRNA has great potential as a new type of antiviral small molecule. In this study, sRNA sequencing and transcriptome sequencing were utilized to investigate alterations in miRNA and mRNA levels in tobacco root, stem, and leaf tissues following PVY infection. The research findings reveal the interplay between miRNAs, PVY, and plant endogenous genes, enhancing our comprehension of the host-PVY virus relationship in diverse tissues. Additionally, this study offers novel insights into the potential employment of miRNAs for viral control.

2 Results

2.1 Analysis of sRNA expression profiles in roots, stems, and leaves of PVY-infected *Nicotiana benthamiana*

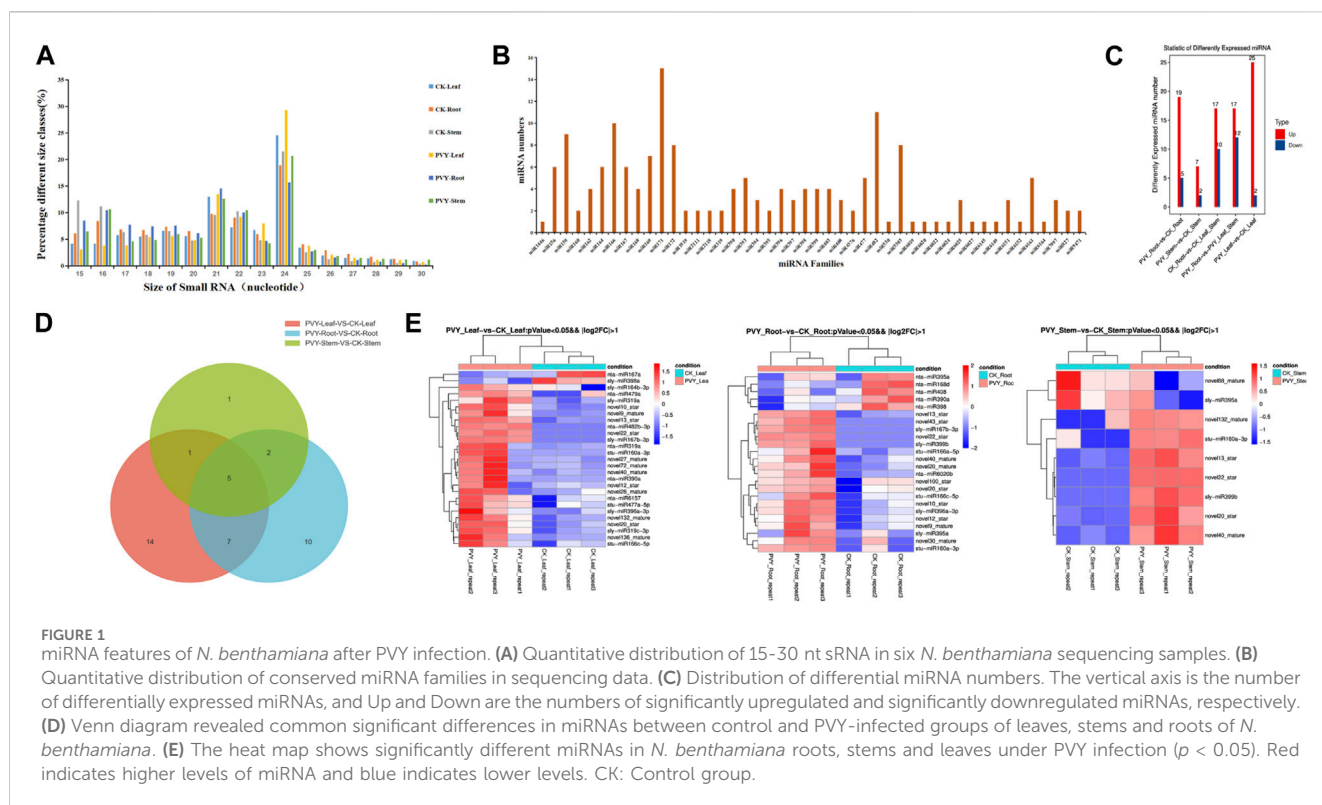
Numerous studies have elucidated the physiological and biochemical changes that occur in plants following viral infection, as plants mount a defense response against viral invasion. sRNAs, particularly miRNAs, have garnered significant attention for their participation in various regulatory pathways governing plant development processes (Umbach and Cullen, 2009). To investigate the alterations in sRNA profiles in plants upon PVY infection, with a focus on miRNAs associated with PVY replication and movement, we generated 18 sRNA libraries derived from PVY-infected *N. benthamiana* (Goodin et al., 2008) roots, stems, and leaves. Supplementary Table S1 provides a summary of the quantities of raw and valid sRNA reads obtained from nine control and nine treatment samples. The results showed that small RNA sizes ranged from 15 to 30 nt, with the most abundant being 21–24 nt as shown in Figure 1A. Following PVY infection, the abundance of sRNAs exhibited diverse changes across the roots, stems, and leaves, with notable variations observed in the size range of 21–24 nt (Figure 1A). The effective reads in the 20–24 nt size category accounted for over 50% of all sRNAs in each sample, with 21 nt and 24 nt reads representing the predominant sRNA types, consistent with findings in dicotyledonous plants (Guo et al., 2017). Notably, we detected a total of 46 miRNA families, although the

number of miRNAs varied within each family. The miR171 family exhibited the highest number of miRNAs (15 members), followed by the miR482 family (11 members), with one-third of the miRNA families comprising only one miRNA member (Figure 1B). Among the 46 miRNA families, we identified 110 known miRNA classes and 141 novel miRNA sequences (Supplementary Table S2).

2.2 Response pattern of miRNAs in *N. benthamiana* upon PVY infection

To comprehensively investigate the alterations in miRNAs expression in *N. benthamiana* following PVY infection and potentially involved in virus infection of miRNA, we conducted an analysis based on normalized reads obtained miRNAs from high-throughput sequencing. The miRBase database (version 22.0) comprising a vast collection of miRNA mature and hairpin precursor sequences from both plants and animals, was utilized for comparisons. Differential expression analysis was performed by comparing the sequencing data with the miRBase database to identify changes in miRNA expression across the 18 sRNA libraries. In comparison to the control group, the majority (73.28%) of differentially expressed miRNAs showed upregulation in PVY-treated samples. Specifically, we identified 25 upregulated and 2 downregulated miRNAs in *N. benthamiana* leaves, including 12 newly discovered miRNAs and 15 conserved miRNAs. In the stems, we identified 6 novel miRNAs and 3 conserved miRNAs, while in the roots, we identified 11 novel miRNAs and 13 conserved miRNAs (Figure 1C). Among the three treatment groups, only 5 miRNAs displayed significant differences across all three tissues. Additionally, 7 miRNAs exhibited significant differences between leaves and roots, 1 miRNA between leaves and stems, and 2 miRNAs between leaves and roots (Figure 1D). Furthermore, when applying more stringent statistical criteria (fold change >2 and p -value ≤ 0.05), we identified 40 miRNAs with significant differences between the PVY treatment and control groups. Among these, 16 miRNAs showed differential expression in at least one pairwise comparison. The expression patterns of these differentially expressed miRNAs are presented in Figure 1E. Notably, we observed inconsistent expression patterns of the same miRNA across the three tissues. For example, miR390a was significantly upregulated in leaves but downregulated in roots, while miR395a showed significant upregulation in leaves and downregulation in stems. This discrepancy suggests that the regulatory targets of miRNAs may vary among different tissues. The varying expression profiles of miRNAs across PVY-infected tissues indicate potential tissue-specific regulatory roles. These findings highlight the involvement of a complex miRNA-mediated regulatory network in PVY infection, underscoring the need for further investigations to elucidate the precise regulatory roles of miRNAs and identify key miRNAs and target genes associated with PVY infection.

sRNAs were extracted from the samples, and the expression levels of the differentially expressed miRNAs identified through sequencing were assessed using qRT-PCR. Figure 2 presents the qRT-PCR detection results, revealing the upregulation of miRNA13, miRNA40, and miRNA20 in all three tissues following PVY infection, with the most prominent increase observed in the leaf,



demonstrating a significant difference. Notably, miRNA10 exhibited significant upregulation in leaves and stems but was undetectable in roots. On the other hand, miRNA398a showed substantial downregulation in both leaves and stems, with a more pronounced reduction observed in the stems. Additionally, the expression levels of miRNA168a and miRNA390a were significantly downregulated in leaves, while their expressions were not detected in roots and stems (Figure 2). Notably, when comparing the qPCR and sequencing results, consistent expression levels of these miRNAs were observed.

2.3 Predictive analysis and functional exploration of potential target genes of miRNAs

To identify key genes involved in miRNA-mediated interactions during PVY infection, we conducted an analysis of the biological functions of target genes corresponding to differentially expressed miRNAs. The *N. benthamiana* genome served as the target database for the differentially expressed miRNAs obtained through sequencing, and targetfinder software was employed for target gene prediction. The analysis revealed that 39 differentially expressed miRNAs (65.00%) exhibited complementarity to 326 potential target sequences, comprising a total of 354 complementary sites. Notably, over 90% of miRNAs in the miRNA-target gene pairs were predicted to target multiple genes, ranging from 2 to 66, while six miRNAs were predicted to target only one gene (Supplementary Table S3). These miRNA-target gene pairs suggest distinct roles for these miRNAs during PVY infection. Figure 3A displays the diverse array of predicted target genes,

encompassing various physiological processes such as growth and development, plant hormones, resistance-related genes, RNA transcription processes, protein-coding genes, enzymes involved in carbohydrate metabolism, and receptor kinases involved in signal transduction (Supplementary Table S4). Among the predicted target genes, miR43 was found to target Auxin-responsive protein IAA29. Additionally, miR10 was implicated in regulating the jasmonate ZIM domain-containing protein (JAZ), thereby modulating the jasmonate signaling pathway. Moreover, miR390a, miR19a, and miR132 were found to participate in stress response by regulating Growth-regulating factor 3, Auxin response factor 18, and Transcriptional regulator TAC1, respectively. In addition, several receptor kinase genes were identified as target genes for specific miRNAs. For instance, miR390a, miR479a, miR477a, and miR168d were found to target LRR receptor-like serine/threonine-protein kinase (RPK2), Serine/threonine-protein kinase (PPK30), Probable serine/threonine protein kinase IREH1, and Probable serine/threonine-protein kinase PBL23, respectively, indicating the regulatory role of miRNAs in receptor kinases during stress responses. Furthermore, target genes related to the virus were also discovered. For example, miR482 and miR408 were found to target TMV resistance protein N (N), which confers resistance to TMV by mediating defense responses that restrict virus replication and movement (Erickson et al., 1999). Moreover, miR168d was found to target heat shock protein 90-5 (HSP90-5), which is crucial for viral homeostasis and promotes viral replication (Geller et al., 2012). These findings suggest the involvement of these miRNA-target pairs in the viral infection process.

To obtain functional annotations for the differentially expressed miRNAs during PVY infection, we conducted a Gene Ontology (GO) classification analysis on the predicted miRNA targets. Out

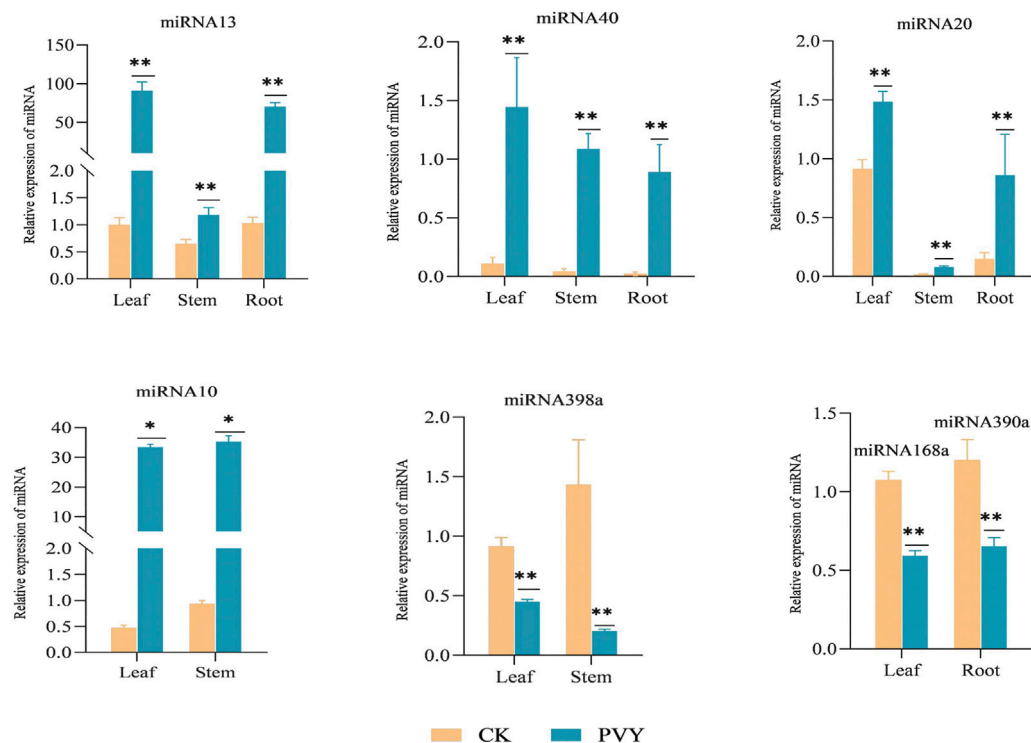


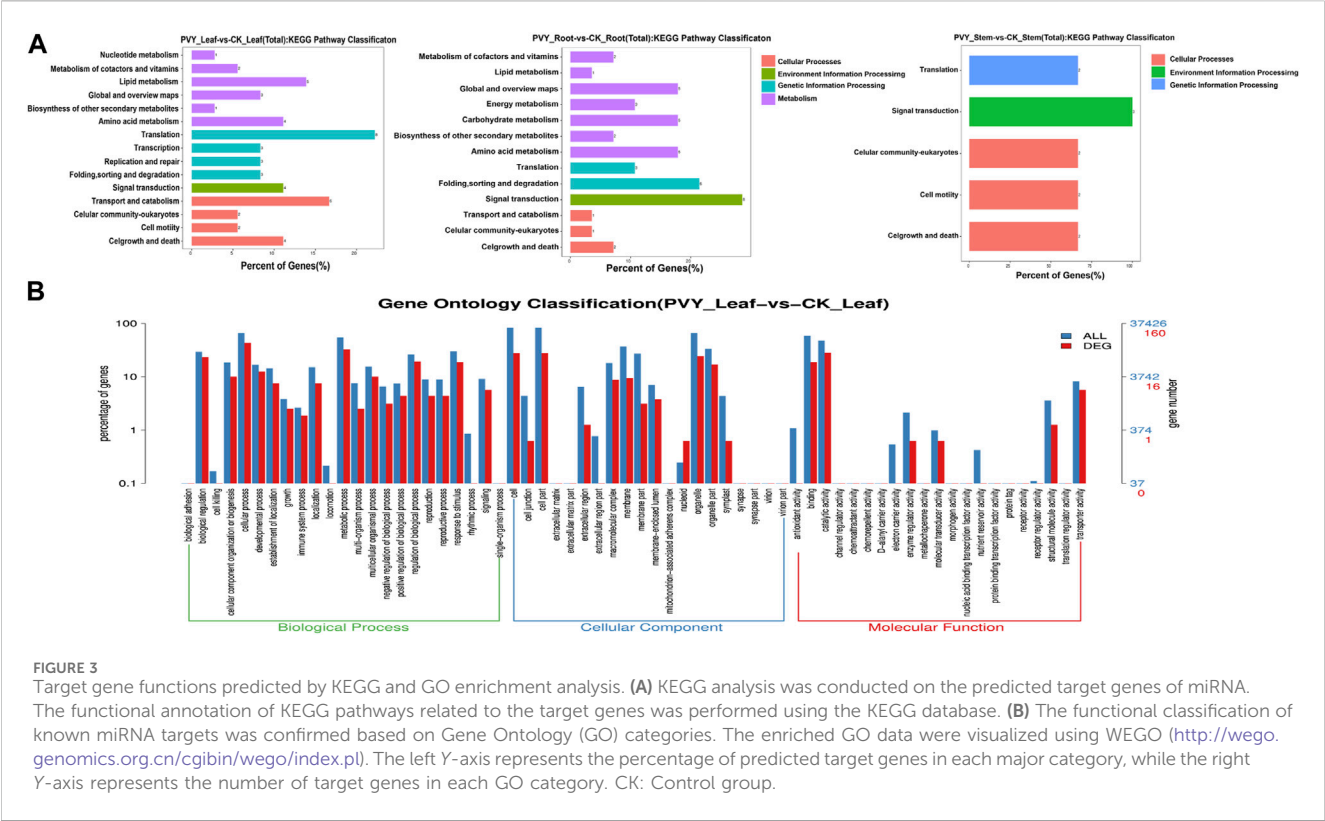
FIGURE 2 qRT-PCR experiments are conducted to confirm the consistency of the expression levels of specific miRNAs, in line with the sequencing results. The figure displays the relative expression of miRNAs in different tissues of the root, stem, and leaf. The yellow bar represents the healthy *N. benthamiana*, while the blue bar represents the *N. benthamiana* after PVY infection. The expression levels of these miRNAs were determined using stem-loop qRT-PCR, with U6 serving as the loading control. The data shown are the mean plus SD of three biological replicates ($n = 3$). Statistical significance was determined using Student's t-test, with $*p < 0.05$ and $**p < 0.01$. CK: Control group.

of the 326 targets associated with the 39 differential miRNAs, 238 targets were found to possess GO terms (Supplementary Table S5). Among the Biological Process classification, the most prominent categories were cellular process, metabolic process, and biological regulation. Additionally, we performed an analysis of the enriched GO terms among the target genes regulated by the differentially expressed miRNAs. As shown in Supplementary Figure S1, the abundance distribution of the three types exhibited similarities to the patterns observed in Figure 3. Specifically, the leaves and roots displayed similar abundance distributions, while the stems exhibited a relatively lower enrichment of physiological processes.

2.4 Integrative analysis of sRNA sequencing and transcriptome sequencing

To explore the interaction between miRNAs and target genes, transcriptome sequencing analysis was conducted on the 18 samples (Mock-inoculated *N. benthamiana* roots, stems and leaves and PVY-infected *N. benthamiana* roots, stems and leaves, three biological replicates for each sample, a total of 18 samples). A total of 7,506 differentially expressed mRNAs were detected (Figure 4A). Comparison of the predicted target genes with the mRNAs detected in the transcriptome revealed that 245 (87.64%) of the predicted target genes were present in the sequencing data.

We compared the predicted target genes of differentially expressed miRNAs detected by sRNA sequencing and the differentially expressed mRNAs detected by transcriptome, and displayed them with a Venn diagram. Further analysis involved comparing the predicted target genes with the differentially expressed mRNAs, resulting in the identification of 36 (13.76%) target genes that matched with the differentially expressed mRNAs (Figure 4B). To gain a more comprehensive understanding of the miRNA-mRNA regulatory network, we constructed a visual representation of the interactions. Notably, the majority of miRNA targets were identified in leaves, with miR319a targeting eight genes, including Phox, SKD1, BIM2, TCP24, TCP2, TCP4, FEP, and SULTR3. Among these, TCP24, TCP4, and BIM2 were targeted simultaneously by multiple miRNAs (Figure 4C). Transcription factors (TFs) play crucial roles in plant defense mechanisms and innate immunity through interactions with cofactors and cisacting elements. The transcriptional regulation of certain TFs can enhance plant defenses against viral infections (Viswanath et al., 2023). Therefore, it is plausible that miRNAs may regulate the viral infection process by targeting TFs. To comprehensively analyze the regulatory network of miRNAs targeting TFs, we performed prediction analysis of transcription factor target genes. The results revealed that the transcription factor ERF had the highest prediction of 18 and 26 target genes in leaves and roots, respectively, while the transcription factor HD-ZIP showed the



highest prediction of 23 target genes in stems (Supplementary Figure S2). These variations may arise from the differential distribution of the virus among the three tissues.

Furthermore, to gain deeper insights into the involvement of target genes in physiological processes, we conducted GO and KEGG enrichment analyses using the sequencing data. We observed that the Cellular Component category exhibited the highest level of enrichment in roots, stems, and leaves, whereas the Molecular Function category displayed the least enrichment. Notably, the enrichment patterns across the three tissues were quite similar (Supplementary Figure S3).

2.5 Predictive analysis of miRNAs targeting the Potato virus Y genome

PVY, a positive-strand RNA virus, belongs to the Potyvirus genus within the Potyviridae family and possesses a genome length of approximately 9.7 kb. Its genome encodes a large polyprotein that can undergo proteolytic cleavage by virus-specific proteases, resulting in the generation of 10 mature proteins (Karasev and Gray, 2013; Quenouille et al., 2013). These proteins include P1, HC-Pro, P3, 6K1, CI, 6K2, VPg, NIa, NIb, and CP. Previous studies have revealed that miRNAs can exert regulatory effects on virus infection by indirectly modulating key genes involved in physiological processes. Additionally, miRNAs can directly target specific sequences within the viral genome, thereby inhibiting viral gene expression. In light of this, we performed target gene

prediction analysis on the differentially expressed miRNAs obtained from the sequencing results and the PVY genome sequence. Our findings revealed that 12 miRNAs were predicted to target specific regions within the PVY genome. Notably, P1 was targeted by miR479a, HC-Pro was targeted by miR482b-3p, P3 was targeted by miR479, anovel10, and miR390a, CI was targeted by miR398a and miR395a, VPg was targeted by miR166a-5p, novel12, novel88, and miR166a-5p, NIa was targeted by novel12, novel88, miR166a-5p, and novel132, NIb was targeted by novel12, miR398a, miR395a, miR6020b, miR6157, and novel30, and CP was targeted by novel88, miR398a, novel43, miR408, and novel22. Remarkably, 8 out of the 10 PVY proteins were found to be targeted by different miRNAs (Table 1). The discovery of miRNAs directly targeting PVY not only aids in the identification of antiviral miRNAs but also offers novel insights for antiviral research.

2.6 Validation of the effects of miR390, miR168, miR32, and miR43 on PVY infection through transient overexpression

During virus infection, miRNAs exert their regulatory effects by targeting key genes involved in various physiological processes, thereby influencing the progression of the viral infection. To explore the specific role of miRNAs in the context of PVY infection, we selected four differentially expressed miRNAs, namely, miR390, miR168, miR32, and miR43, based on their predicted target genes and functional

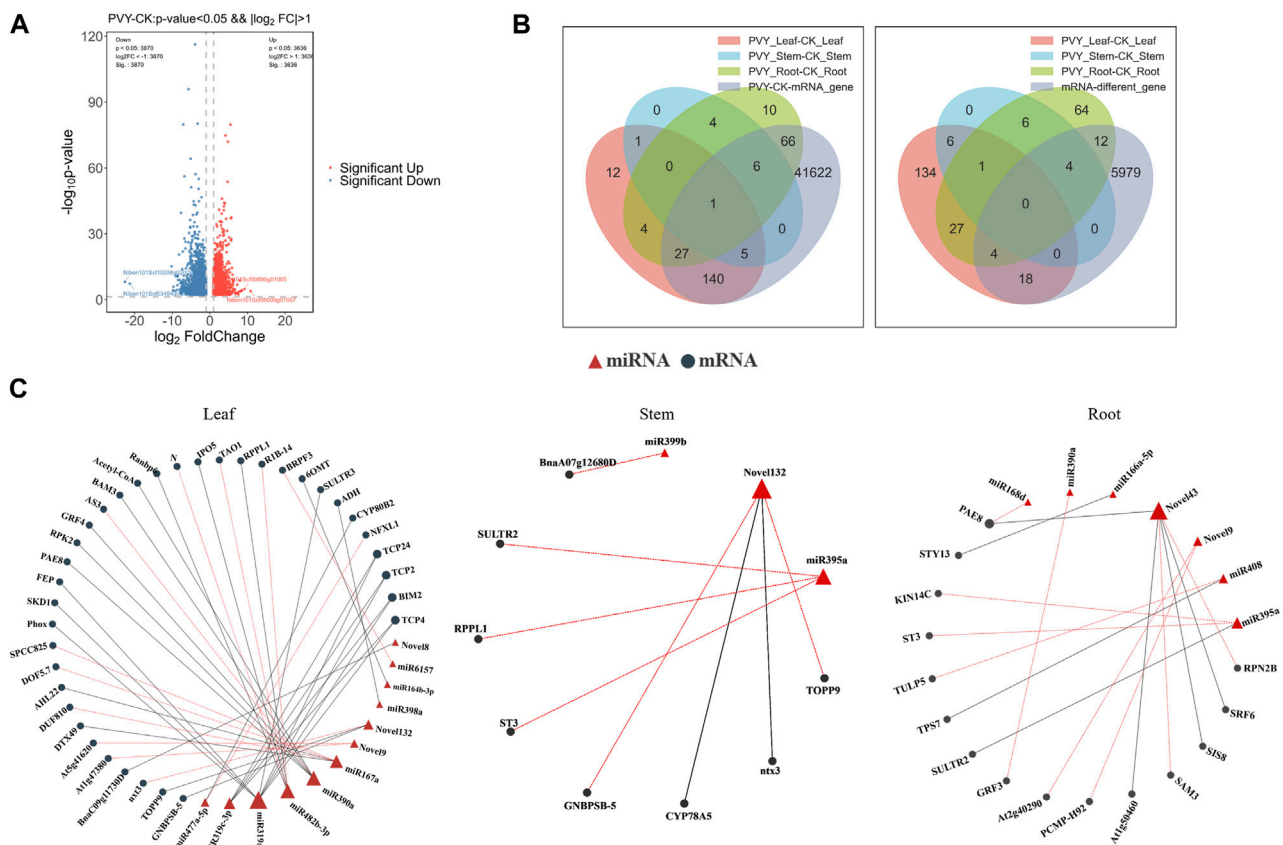


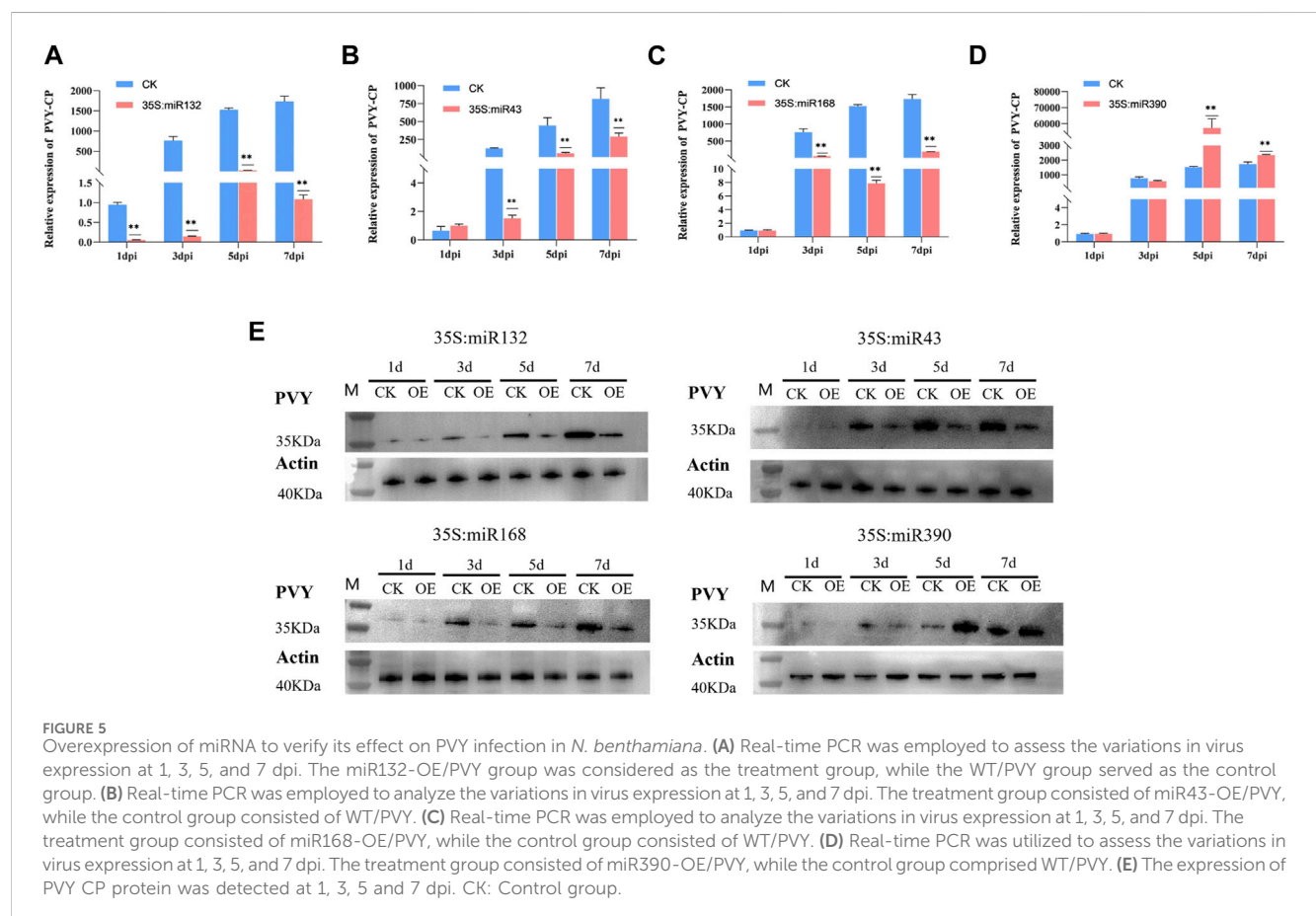
FIGURE 4 Combined analysis of small RNA sequencing and transcriptome sequencing further clarified the existing miRNA-mRNA pairs and their functions. **(A)** Transcriptome differential gene volcano map. The difference generated by the comparison is reflected in the volcano map, the gray is the non-significant difference gene, and the red and blue are the significant difference gene; the horizontal axis is \log_2 FoldChange, and the vertical axis is $\log_{10}q\text{-value}$. **(B)** Venn diagram showing overlap of miRNA target genes with transcriptome total and differentially expressed mRNAs. The left side shows the distribution of the number of differentially expressed mRNAs in roots, stems and leaves, as well as the total differentially expressed mRNAs. The overlapping parts represent shared mRNAs; the right side shows the comparison of the predicted target genes of miRNAs and the differentially expressed mRNAs in the transcriptome. **(C)** miRNA-mRNA network diagram. The relationship between miRNAs and target genes, red triangles represent miRNAs, black circles represent target genes, red lines represent upregulation of target genes, and black lines represent downregulation of target genes. CK: Control group.

analysis outcomes. The miR390 family is widely distributed among various plants and plays a role in regulating plant growth and development processes. However, there have been limited studies on its involvement in biotic stress. Previous research has shown that miR168 is involved in the infection process of various viruses. Our attention was drawn to the prediction results of miR43, which targets multiple regulatory genes including AUX/IAA. Additionally, the results of KEGG analysis revealed that miR132 in the stem targets multiple pathways such as MAPK and PI3K-Akt, which we chose to focus on. Overexpression vectors carrying miR390, miR168, miR132, and miR43 were constructed and transformed into *Agrobacterium* strain LBA4404, followed by introduction into *N. benthamiana* plants previously inoculated with PVY via *Agrobacterium*-mediated method. In addition, the transcript and protein levels of PVY coat protein (CP) were examined using qRT-PCR and Western blotting, respectively. qRT-PCR was employed to assess the virus expression levels in both the control and treatment groups at 1,3,5 and 7 days post-infection

(dpi). The results revealed that overexpression of miR132, miR43, and miR168 significantly reduced the transcript levels of PVY-CP in the treatment group compared to the control group at 3 dpi, 5 dpi, and 7 dpi after PVY infection. Notably, miR132 exhibited the most pronounced inhibitory effect on the virus, with inhibition rates exceeding 50% following PVY inoculation. Moreover, the expression of CP protein in the control and treatment groups of *N. benthamiana* showed significant differences (Figures 5A–C). Conversely, plants overexpressing miR390 exhibited significantly increased transcript levels of PVY-CP at 5 dpi and 7 dpi compared to the control group. Notably, the virus increased by more than 10-fold at 5 dpi, and the expression of CP protein in the PVY control and treatment groups displayed notable distinctions (Figure 5D). In this study, we employed PVY-GFP to infect *N. benthamiana* plants that overexpressed four types of miRNA (Figure 6). After 7 days, we observed the fluorescence intensity and phenotype, which were found to be consistent with the Western blotting results. These results demonstrate the regulatory role of miRNAs

TABLE 1 High-confidence binding sites of miRNAs targeting the PVY genome were predicted by the miranda algorithm.

miRNA	Total energy	Max energy	Positions	miRNA length	Total score
miR479a	-42.58	-17.13	641 3115 2485	22	472
novel88	-40.11	-15.59	6584 6186 8938	19	440
miR398a	-40.11	-16.04	9409 5485 7698	21	439
miR166a-5p	-25.93	-15.68	7062 6512	21	309
novel43	-38.56	-20.29	9322 9535	21	298
miR395a	-29.69	-16.11	3873 7167	21	286
miR6020b	-15.4	-15.4	7700	21	172
miR408	-27.75	-27.75	8901	21	164
miR6157	-20.49	-20.49	7763	22	156
miR482b-3p	-15.01	-15.01	1563	22	154
novel30	-19.5	-19.5	8341	20	152
miR390a	-17.3	-17.3	2473	21	141



in modulating PVY infection. To further assess the expression of PVY-CP protein, Western blotting analysis was performed. The results were consistent with the qRT-PCR results, indicating that overexpression of miR132, miR43, and miR168 inhibited PVY

infection, whereas overexpression of miR390 promoted viral infection. Taken together, these findings highlight the regulatory capacity of miRNAs in PVY infection and underline their significance in plant antiviral immunity.

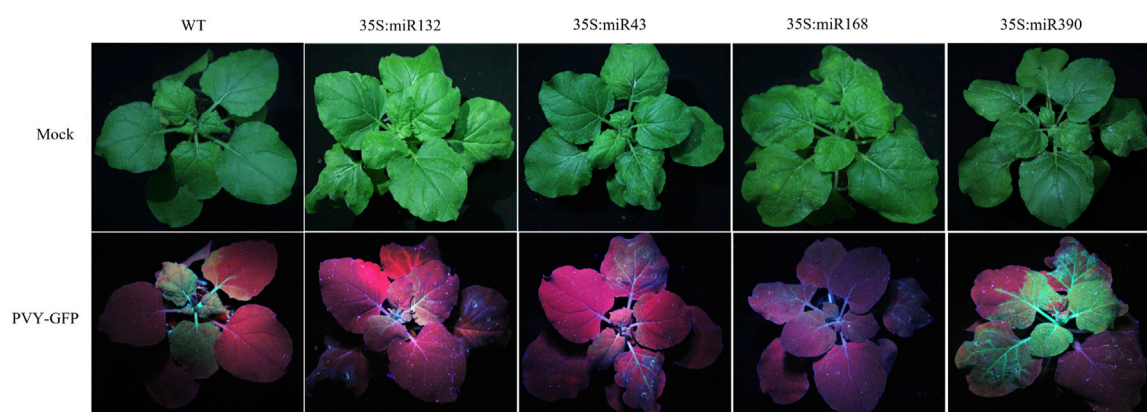


FIGURE 6
PVY-GFP was used to infect *N. benthamiana* treated with different overexpression of miRNA, and the degree of PVY infection was observed using UV light.

3 Discussion

Over the past decade, advancements in high-throughput technologies have greatly facilitated genome and transcriptome sequencing. As a result, there has been a continuous expansion of plant genome and transcriptome databases, which has allowed for the comprehensive identification and functional annotation of plant miRNAs and siRNAs. Several studies have demonstrated that certain miRNAs can play a role in virus infection and enhance a plant's ability to resist viruses (Lu et al., 2008; Prasad et al., 2019). In this study, we conducted sRNA sequencing and transcriptome sequencing on *N. benthamiana* after PVY infection, and performed a joint analysis of the obtained sequencing results. A total of 60 differentially expressed miRNAs were identified, with 51 upregulated and 9 downregulated. Among these differentially expressed miRNAs, miR168 (Ludman and Fatyol, 2021), miR167 (Liu et al., 2022), miR319 (Zhang et al., 2016), miR164 (Bazzini et al., 2009), miR166 (Prasad et al., 2022), and miR398 (Lin et al., 2022) have been previously associated with virus infection. KEGG enrichment analysis revealed that the predicted targets of miRNAs in the roots, stems, and leaves showed different major enriched signaling pathways. The roots were mainly enriched in PI3K-Akt (Ji and Liu, 2008), Apoptosis (Everett and McFadden, 1999), Wnt (Rim et al., 2022), and Plant hormone (Blazquez et al., 2020) signaling pathways, while the leaves and stems were mainly enriched in RAS, cGMP-PKG, MAPK (Kumar et al., 2018) and other signaling pathways. These signaling pathways are related to viral infection and replication, and the different tissues result in different miRNAs enriching different signaling pathways.

The expression profiles of miRNAs were analyzed after PVY infection in different tissues, revealing inconsistent expression patterns. This inconsistency may be attributed to the ability of miRNAs to target multiple genes and the varying effects of the virus on different tissue physiological processes (Figure 1). By utilizing sRNA and transcriptome sequencing, a potential miRNA-mRNA regulatory network was constructed, as depicted in Figure 4. Through the prediction of miRNA-PVY genome

interactions, it was identified that 12 miRNAs target 21 specific sites within the PVY genome (Table 1). The regulatory network revealed the involvement of multiple immune pathways, including the targeting of the tobacco resistance N gene by miR482, which confers tobacco TMV resistance (Dinesh-Kumar and Baker, 2000). Additionally, phytohormones play a crucial role in regulating plant defense responses to various biotic and abiotic stresses, as well as plant growth and development (Huot et al., 2014). This study also discovered several miRNAs that target various plant hormones. MiR43 was predicted to target the key gene AUX/IAA in the auxin pathway, while miR10 targeted the JAZ gene in the JA pathway (jasmonate ZIM domain-containing protein). JAZ is involved in the JA signaling pathway, which regulates plant defense responses (Huang et al., 2022). To evaluate the impact of specific miRNAs on PVY infection, overexpression experiments were conducted on four miRNAs. The results demonstrated that miR168, miR132, and miR43 effectively inhibited virus infection, while miR390 promoted PVY infection (Figure 5). In our study, we discovered that miR168 targets HSP90, which plays a crucial role in viral replication (Verchot, 2012). Additionally, we observed a downregulation of miR168 expression and an increase in HSP90 mRNA expression following PVY infection. Based on these findings, we hypothesized that miR168 targets HSP90 to delay virus replication. Another miRNA, miR390, is predicted to target ACAA1 (acetyl-CoA acyltransferase 1), which is potentially involved in the JA biosynthetic pathway. Previous studies have suggested that miR390 indirectly influences viral infection by regulating plant hormones such as JA and SA (Oka et al., 2013). These studies collectively indicate that miRNA168 and miR390 regulate viral infection through different mechanisms.

Overall, plant-pathogen interactions result in a delicate balance between the plant immune signaling network and the pathogens it encounters. In this study, our focus was on exploring the potential miRNA-mRNA network after PVY infection. We selected four miRNAs for overexpression to assess their impact on PVY infection. This research provides a foundation for the development of effective anti-PVY miRNAs. Our findings

demonstrate that the miRNA regulatory network directly affects PVY infection by targeting resistance genes or key genes involved in PVY replication. Moreover, it can also indirectly regulate viral infection by targeting key genes in immune and developmental pathways.

4 Materials and methods

4.1 Plant materials and viral strains

N. benthamiana plants were cultivated in a greenhouse under controlled conditions, with a photoperiod of 16/8 h (light/dark) and a temperature of 25°C. Healthy *N. benthamiana* plants at the 4-week-old stage were subjected to PVY infection or mock inoculation, and tissue samples were collected 7 days post-treatment. Three biological replicates were obtained for each sample. For PVY infection, the infected PVY leaves were diluted 40-fold with phosphate-buffered saline (PBS) and subsequently inoculated onto *N. benthamiana* leaves. Control group (CK): Use 1XPBS to simulate infection on healthy plants. PVY treatment: Use the preserved PVYN:O or PVY-GFP strain to infect *N. benthamiana* (Sun et al., 2018).

4.2 Sample preparation

The treatments for *N. benthamiana* are as follows: the control group: using PBS for simulated treatment; the treatment group: using PVY virus to treat the leaves through quartz sand. Sampling was carried out 7 days after treatment, and the roots, stems and leaves of the control group (CK-Root-Repeat1, CK-Root-Repeat2, CK-Root-Repeat3, CK-Stem-Repeat1, CK-Stem-Repeat2, CK-Stem-Repeat3, CK-Leaf-Repeat1, CK-Leaf-Repeat2, CK-Leaf-Repeat3) and the roots, stems and leaves of the treatment group (PVY-Root-Repeat1, PVY-Root-Repeat2, PVY-Root-Repeat3, PVY-Stem-Repeat1, PVY-Stem-Repeat2, PVY-Stem-Repeat3, PVY-Leaf-Repeat1, PVY-Leaf-Repeat2, PVY-Leaf-Repeat3), three biological replicates for each tissue, a total of 18 samples. Then perform sRNA extraction and construct sRNA library.

4.3 sRNA library construction and sequencing

sRNAs were extracted from the samples using the Illumina TruSeq Small RNA Preparation Kit (Illumina, San Diego, CA, United States) to construct sRNA libraries. The sRNA library construction involved ligating 10 µg of total RNA from the 18 samples with 3' and 5' adapters using T4 RNA ligase. After amplification, 1 µg of the PCR product was subjected to sRNA sequencing on the Agilent 2100 chip platform at Shanghai Ouyi Biotechnology (Shanghai, China) (Billmeier and Xu, 2017). The length and quality of the resulting libraries were confirmed based on the obtained sequencing results. To identify miRNAs within the data, adapter sequences were initially removed from the original reads using cutadapt (Martin, 2011), and sequences were filtered based on length, excluding those shorter than 15 bp or longer than

41 bp. Subsequently, NGSQCToolkit (version 2.3.2) (Patel and Jain, 2012) was employed to eliminate reads containing N bases, resulting in high-quality clean reads. All valid reads from the 18 sRNA libraries were then aligned to the *N. benthamiana* genome and known plant miRNA sequences. Unannotated sRNA sequences were subjected to prediction using Mirdeep2 (Friedlander et al., 2012) software to identify novel miRNAs, and the secondary structures of novel miRNAs were predicted using RNAfold (Hofacker, 2003) software.

4.4 Bioinformatics analysis of sRNA sequences

4.4.1 miRNA expression analysis

Raw reads underwent adapter removal using cutadapt (Kechin et al., 2017), followed by filtering based on length to discard sequences shorter than 15 bp or longer than 41 bp. Quality control was performed using fastx toolkit (version 0.0.13) (http://hannonlab.cshl.edu/fastx_toolkit) software, retaining sequences with a Q20 score of 80% or higher. NGSQCToolkit (version 2.3.2) (Patel and Jain, 2012) was then employed to filter out reads containing N bases, resulting in a collection of high-quality clean reads. The length distribution of these clean reads was statistically analyzed to gain preliminary insights into the distribution of sRNAs within the samples. To determine the proportion of reads aligned to the genome, clean reads were aligned against the reference genome sequence of *N. benthamiana*. Subsequently, clean reads were aligned to the Rfam database using blastn software, enabling annotation of rRNA, snRNA, snoRNA, tRNA, and other sequences. An additional filtering step removed these annotated sequences based on the Rfam database. Moreover, reads matching the transcript sequences with a length less than 15 bp or greater than 41 bp were excluded. RepeatMasker (Chen, 2004) software was utilized to compare the filtered sequences with the repeat database, leading to the identification and removal of repetitive sequences.

4.4.2 miRNA differential analysis

Expression statistics were performed based on the sequences of both known mature miRNAs and the newly predicted miRNAs, with miRNA expression calculated using TPM (transcripts per million) as the metric index. Specifically, TPM (Sun et al., 2014) was determined as the number of reads aligning to each miRNA per sample divided by the total number of aligned reads, multiplied by 106. The significance of differential expression was evaluated using the DEG algorithm within the R package, with miRNAs exhibiting a *p*-value <0.05 considered as differentially expressed.

4.4.3 Differential miRNA target gene prediction and target gene function analysis

Target genes were predicted using miranda and annotated in the reference genome. Target gene prediction for differentially expressed known miRNAs and newly predicted miRNAs was conducted using the targetfinder software. The hypergeometric distribution test of R software was employed for KEGG and GO analyses of the target genes, with the *p*-values subsequently corrected by Benjamini and Hochberg's multiple tests to obtain the false discovery rate (FDR).

4.4.4 Validation of differentially expressed miRNAs and virus using qRT-PCR

The expression levels of miRNAs identified through sRNA sequencing were verified by qRT-PCR. For expression analysis, we selected six differentially expressed miRNAs and performed qRT-PCR using the 7500 Fast Real-Time PCR System (Applied Biosystems). The forward primer is designed and synthesized based on the specific miRNA sequence, while the reverse primer uses the universal primer Q in the kit (Supplementary Table S6). miRNA extraction from the 18 samples was performed using the EasyPure miRNA Kit (TransGen Biotech, Beijing, China). Subsequently, the extracted RNA underwent reverse transcription using the miR first Strand cDNA synthesis kit (Vazyme, Nanjing, China). The qRT-PCR reaction was conducted on the ABI7500 platform using the $2^{-\Delta\Delta CT}$ method to calculate the relative expression levels of miRNAs. U6 and actin were utilized as endogenous control and control group1 standard Ct values, respectively. Each experiment consisted of three biological and technical replicates (Song et al., 2023).

4.4.5 Preparation of miRNA expression vectors

To construct the pGate-100-miRNA vector, the miRNA sequence was inserted into the mature miR319a precursor using pRS300 plasmid and specific primers (Ossowski et al., 2008). The miR319a-containing miRNA sequence was then inserted into the pGate-100 expression vector harboring the 35S promoter. The recombinant plasmids were sequenced to confirm correct insertion of the miRNA backbone sequence. Recombinant plasmids were subsequently retransformed into *Escherichia coli* and confirmed by PCR through single colony selection. Purification of recombinant plasmids was accomplished using the Rapid Pure Endogenous Plasmid Maxi Kit (DC202-01; Vazyme, Nanjing, China). Plasmid concentration, average optical density (OD) A260/A280, and plasmid purity QA/QC were determined. The target gene was amplified by PCR using pGate-100 universal primers, and the PCR products were subjected to agarose gel electrophoresis, followed by recovery and storage at -20°C .

4.4.6 Western blotting assay

Total protein was extracted from the samples using the Pierce Classic IP kit (Pierce), and protein concentration was determined using the BCA working solution. The protein of interest was separated on a 10% SDS polyacrylamide gel. Immunoblot analysis of PVY-CP was conducted using a rabbit antibody (1:5000, Invitrogen), while viral CP protein was detected using a PVY-CP antibody (1:2000) (Song et al., 2023). The antigens were detected by chemiluminescence using the ECL reagent (SuperSignal West Pico kit) after transmembrane transfer. The membrane was then subjected to fluorescence detection using a fluorescence imager, with β -actin serving as an internal reference for quantitative analysis of band intensity. Three independent experiments were performed for analysis.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and

accession number(s) can be found below: <https://www.ncbi.nlm.nih.gov/>, PRJNA999491; <https://www.ncbi.nlm.nih.gov/>, PRJNA991318.

Author contributions

HS: Formal Analysis, Investigation, Visualization, Writing—original draft, Writing—review and editing. XG: Data curation, Software, Validation, Visualization, Writing—original draft. LSo: Methodology, Writing—review and editing. YJ: Resources, Writing—review and editing. LSh: Data curation, Writing—review and editing. JY: Project administration, Writing—review and editing. CL: Supervision, Writing—review and editing. JS: Writing—review and editing, Funding acquisition. HW: Project administration, Writing—review and editing. SZ: Conceptualization, Writing—review and editing. YL: Conceptualization, Writing—review and editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the Major science and technology projects: [110202201020 (LS-04), 2022JBXM03, 2023410300200041]; Shandong Provincial Natural Science Foundation Project (ZR2021MC058); Qingdao Science and Technology Projects (23-2-8-xdny-12-nsh).

Conflict of interest

Authors CL and JS were employed by Liupanshui City Company of Guizhou Tobacco Company.

Author HW was employed by Luoyang City Company of Henan Tobacco Company.

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/fgene.2023.1290466/full#supplementary-material>

SUPPLEMENTARY TABLE S1

sRNA raw read information. The raw image data files obtained by high-throughput sequencing are converted into raw sequencing sequence

RawData through base calling analysis; reads containing N bases are filtered out. Finally, high-quality CleanReads are obtained; the number of categories of CleanReads is counted and uniq reads are obtained. CleanReads performs length distribution statistics and counts the repetition frequency of uniq reads.

SUPPLEMENTARY TABLE S2

Among the 46 miRNA families, we identified 110 known miRNA classes and 141 novel miRNA sequences. The number and classification of miRNA families detected in the tissues of tobacco roots, stems and leaves under control treatment and PVY infection; new miRNA sequences detected in sRNA sequencing.

SUPPLEMENTARY TABLE S3

Predicted target gene information of differentially expressed miRNAs in roots, stems and leaves.

SUPPLEMENTARY TABLE S4

Predicted target gene information. Annotation and detailed description of differential miRNAs in roots, stems and leaves in the predicted target genes of *Nicotiana benthamiana* reference genome.

SUPPLEMENTARY TABLE S5

GO enrichment classification information of target genes predicted by differential miRNAs. GO functional analysis is to functionally annotate and classify all genes and differential miRNA target genes of the species. It includes three major categories: Biological Process (BP) represents the process of a molecular activity event; Cellular Component (CC) represents the cell or its external environment; Molecular Function (MF) It is the active element that describes the gene product at the molecular level.

References

- Ai, T., Zhang, L., Gao, Z., Zhu, C. X., and Guo, X. (2011). Highly efficient virus resistance mediated by artificial microRNAs that target the suppressor of pvx and pvv in plants. *Plant Biol.* 13 (2), 304–316. doi:10.1111/j.1438-8677.2010.00374.x
- Bazzini, A. A., Almasia, N. I., Manacorda, C. A., Mongelli, V. C., Conti, G., Maroniche, G. A., et al. (2009). Virus infection elevates transcriptional activity of mir164a promoter in plants. *BMC Plant Biol.* 9, 152. doi:10.1186/1471-2229-9-152
- Billmeier, M., and Xu, P. (2017). Small rna profiling by next-generation sequencing using high-definition adapters. *Methods Mol. Biol.* 1580, 45–57. doi:10.1007/978-1-4939-6866-4_4
- Blazquez, M. A., Nelson, D. C., and Weijers, D. (2020). Evolution of plant hormone response pathways. *Annu. Rev. Plant Biol.* 71, 327–353. doi:10.1146/annurev-arplant-050718-100309
- Bologna, N. G., and Voinnet, O. (2014). The diversity, biogenesis, and activities of endogenous silencing small rnas in arabidopsis. *Annu. Rev. Plant Biol.* 65, 473–503. doi:10.1146/annurev-arplant-050213-035728
- Boualem, A., Dogimont, C., and Bendahmane, A. (2016). The battle for survival between viruses and their host plants. *Curr. Opin. Virol.* 17, 32–38. doi:10.1016/j.coviro.2015.12.001
- Chen, N. (2004). Using repeatmasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinforma.* 4, 4.10–10. doi:10.1002/0471250953.bi0410s05
- Dang, Y., Yang, Q., Xue, Z., and Liu, Y. (2011). RNA interference in fungi: pathways, functions, and applications. *Eukaryot. Cell* 10 (9), 1148–1155. doi:10.1128/EC.05109-11
- Davis, S., Lollo, B., Freier, S., and Esau, C. (2006). Improved targeting of miRNA with antisense oligonucleotides. *Nucleic Acids Res.* 34 (8), 2294–2304. doi:10.1093/nar/gkl183
- Dinesh-Kumar, S. P., and Baker, B. J. (2000). Alternatively spliced n resistance gene transcripts: their possible role in tobacco mosaic virus resistance. *Proc. Natl. Acad. Sci. U. S. A.* 97 (4), 1908–1913. doi:10.1073/pnas.020367497
- Ding, S. W., and Voinnet, O. (2007). Antiviral immunity directed by small rnas. *Cell* 130 (3), 413–426. doi:10.1016/j.cell.2007.07.039
- Duan, C. G., Wang, C. H., Fang, R. X., and Guo, H. S. (2008). Artificial microRNAs highly accessible to targets confer efficient virus resistance in plants. *J. Virol.* 82 (22), 11084–11095. doi:10.1128/JVI.01377-08
- Erickson, F. L., Dinesh-Kumar, S. P., Holzberg, S., Ustach, C. V., Dutton, M., Handley, V., et al. (1999). Interactions between tobacco mosaic virus and the tobacco n gene. *Philos. Trans. R. Soc. B-Biol. Sci.* 354 (1383), 653–658. doi:10.1098/rstb.1999.0417
- Everett, H., and McFadden, G. (1999). Apoptosis: an innate immune response to virus infection. *Trends Microbiol.* 7 (4), 160–165. doi:10.1016/s0966-842x(99)01487-0
- Fahim, M., Millar, A. A., Wood, C. C., and Larkin, P. J. (2012). Resistance to wheat streak mosaic virus generated by expression of an artificial polycistronic microRNA in wheat. *Plant Biotechnol. J.* 10 (2), 150–163. doi:10.1111/j.1467-7652.2011.00647.x
- Friedlander, M. R., Mackowiak, S. D., Li, N., Chen, W., and Rajewsky, N. (2012). Mirdeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res.* 40 (1), 37–52. doi:10.1093/nar/gkr688
- Geller, R., Taguwa, S., and Frydman, J. (2012). Broad action of hsp90 as a host chaperone required for viral replication. *Biochim. Biophys. Acta* 1823 (3), 698–706. doi:10.1016/j.bbamcr.2011.11.007
- Ghildiyal, M., and Zamore, P. D. (2009). Small silencing rnas: an expanding universe. *Nat. Rev. Genet.* 10 (2), 94–108. doi:10.1038/nrg2504
- Goodin, M. M., Zaitlin, D., Naidu, R. A., and Lommel, S. A. (2008). *Nicotiana benthamiana*: its history and future as a model for plant-pathogen interactions. *Mol. Plant-Microbe Interact.* 8 (21), 1015–1026. doi:10.1094/MPMI-21-8-1015
- Gray, S., De Boer, S., Lorenzen, J., Karasev, A., Whitworth, J., Nolte, P., et al. (2010). Potato virus Y: an evolving concern for potato crops in the United States and Canada. *PLANT Dis.* 94 (12), 1384–1397. doi:10.1094/PDIS-02-10-0124
- Guo, Y., Zhao, S., Zhu, C., Chang, X., Yue, C., Wang, Z., et al. (2017). Identification of drought-responsive miRNAs and physiological characterization of tea plant (*Camellia sinensis* L.) under drought stress. *BMC Plant Biol.* 17 (1), 211. doi:10.1186/s12870-017-1172-6
- Hofacker, I. L. (2003). Vienna RNA secondary structure server. *Nucleic Acids Res.* 31 (13), 3429–3431. doi:10.1093/nar/gkg599
- Huang, H., Zhao, W., Li, C., Qiao, H., Song, S., Yang, R., et al. (2022). Slvq15 interacts with jasmonate-zim domain proteins and slwrky31 to regulate defense response in tomato. *Plant Physiol.* 190 (1), 828–842. doi:10.1093/plphys/kiac275
- Huot, B., Yao, J., Montgomery, B. L., and He, S. Y. (2014). Growth-defense tradeoffs in plants: a balancing act to optimize fitness. *Mol. Plant.* 7 (8), 1267–1287. doi:10.1093/mp/ssu049
- Ji, W. T., and Liu, H. J. (2008). PI3K-akt signaling and viral infection. *Recent Pat. Biotechnol.* 2 (3), 218–226. doi:10.2174/187220808786241042
- Karasev, A. V., and Gray, S. M. (2013). Continuous and emerging challenges of potato virus Y in potato. *Annu. Rev. Phytopathol.* 51, 571–586. doi:10.1146/annurev-phyto-082712-102332
- Kechin, A., Boyarskikh, U., Kel, A., and Filipenko, M. (2017). Cutprimers: a new tool for accurate cutting of primers from reads of targeted next generation sequencing. *J. Comput. Biol.* 24 (11), 1138–1143. doi:10.1089/cmb.2017.0096

- Kumar, R., Khandelwal, N., Thachamvally, R., Tripathi, B. N., Barua, S., Kashyap, S. K., et al. (2018). Role of mapk/mnk1 signaling in virus replication. *Virus Res.* 253, 48–61. doi:10.1016/j.virusres.2018.05.028
- Li, F., Pignatta, D., Bendix, C., Brunkard, J. O., Cohn, M. M., Tung, J., et al. (2012). MicroRNA regulation of plant innate immune receptors. *Proc. Natl. Acad. Sci. U. S. A.* 109 (5), 1790–1795. doi:10.1073/pnas.1118282109
- Lin, K. Y., Wu, S. Y., Hsu, Y. H., and Lin, N. S. (2022). Mir398-regulated antioxidants contribute to bamboo mosaic virus accumulation and symptom manifestation. *Plant Physiol.* 188 (1), 593–607. doi:10.1093/plphys/kiab451
- Liu, S. R., Zhou, J., Hu, C. G., Wei, C. L., and Zhang, J. Z. (2017a). MicroRNA-mediated gene silencing in plant defense and viral counter-defense. *Front. Microbiol.* 8, 1801. doi:10.3389/fmicb.2017.01801
- Liu, W. W., Meng, J., Cui, J., and Luan, Y. S. (2017b). Characterization and function of microRNA(*)s in plants. *Front. Plant Sci.* 8, 2200. doi:10.3389/fpls.2017.02200
- Liu, X., Liu, S., Chen, X., Prasanna, B. M., Ni, Z., Li, X., et al. (2022). Maize mir167-arf3/30-polyamine oxidase 1 module-regulated h₂o₂ production confers resistance to maize chlorotic mottle virus. *Plant Physiol.* 189 (2), 1065–1082. doi:10.1093/plphys/kiac099
- Lu, Y. D., Gan, Q. H., Chi, X. Y., and Qin, S. (2008). Roles of microRNA in plant defense and virus offense interaction. *Plant Cell Rep.* 27 (10), 1571–1579. doi:10.1007/s00299-008-0584-z
- Ludman, M., and Fatyol, K. (2021). Targeted inactivation of the ago1 homeologues of *Nicotiana benthamiana* reveals their distinct roles in development and antiviral defence. *New Phytol.* 229 (3), 1289–1297. doi:10.1111/nph.16992
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *Emblnet J.* 17 (1), 10–12. doi:10.14806/ej.17.1.200
- Mengistu, A. A., and Tenkegna, T. A. (2021). The role of mirna in plant-virus interaction: a review. *Mol. Biol. Rep.* 48 (3), 2853–2861. doi:10.1007/s11033-021-06290-4
- Meziadi, C., Blanchet, S., Geffroy, V., and Pflieger, S. (2017). Genetic resistance against viruses in *Phaseolus vulgaris* L.: State of the art and future prospects. *Plant Sci.* 265, 39–50. doi:10.1016/j.plantsci.2017.08.009
- Naqvi, A. R., Sarwat, M., Pradhan, B., Choudhury, N. R., Haq, Q. M., and Mukherjee, S. K. (2011). Differential expression analyses of host genes involved in systemic infection of tomato leaf curl new delhi virus (tolcndv). *Virus Res.* 160 (1–2), 395–399. doi:10.1016/j.virusres.2011.05.002
- Navarro, L., Dunoyer, P., Jay, F., Arnold, B., Dharmasiri, N., Estelle, M., et al. (2006). A plant miRNA contributes to antibacterial resistance by repressing auxin signaling. *Science* 312 (5772), 436–439. doi:10.1126/science.1126088
- Niu, Q. W., Lin, S. S., Reyes, J. L., Chen, K. C., Wu, H. W., Yeh, H. D., et al. (2006). Expression of artificial microRNAs in transgenic arabidopsis thaliana confers virus resistance. *Nat. Biotechnol.* 24 (11), 1420–1428. doi:10.1038/nbt1255
- Oka, K., Kobayashi, M., Mitsuhashi, I., and Seo, S. (2013). Jasmonic acid negatively regulates resistance to tobacco mosaic virus in tobacco. *Plant Cell Physiol.* 54 (12), 1999–2010. doi:10.1093/pcp/pct137
- Ossowski, S., Schwab, R., and Weigel, D. (2008). Gene silencing in plants using artificial microRNAs and other small RNAs. *Plant J.* 53 (4), 674–690. doi:10.1111/j.1365-3113.2007.03328.x
- Patel, R. K., and Jain, M. (2012). Ngs qc toolkit: a toolkit for quality control of next generation sequencing data. *PLoS One* 7 (2), e30619. doi:10.1371/journal.pone.0030619
- Prasad, A., Sharma, N., Chirom, O., and Prasad, M. (2022). The sly-mir166-slyhb module acts as a susceptibility factor during tolcnv infection. *Theor. Appl. Genet.* 135 (1), 233–242. doi:10.1007/s00122-021-03962-4
- Prasad, A., Sharma, N., Muthamilarasan, M., Rana, S., and Prasad, M. (2019). Recent advances in small RNA mediated plant-virus interactions. *Crit. Rev. Biotechnol.* 39 (4), 587–601. doi:10.1080/07388551.2019.1597830
- Qin, L., Ding, S., and He, Z. (2023). Compositional biases and evolution of the largest plant RNA virus order potatavirales. *Int. J. Biol. Macromol.* 240, 124403. doi:10.1016/j.ijbiomac.2023.124403
- Quenouille, J., Vassilakos, N., and Moury, B. (2013). Potato virus Y: a major crop pathogen that has provided major insights into the evolution of viral pathogenicity. *Mol. Plant Pathol.* 14 (5), 439–452. doi:10.1111/mpp.12024
- Rim, E. Y., Clevers, H., and Nusse, R. (2022). The Wnt pathway: from signaling mechanisms to synthetic modulators. *Annu. Rev. Biochem.* 91, 571–598. doi:10.1146/annurev-biochem-040320-103615
- Shivaprasad, P. V., Chen, H. M., Patel, K., Bond, D. M., Santos, B. A., and Baulcombe, D. C. (2012). A microRNA superfamily regulates nucleotide binding site-leucine-rich repeats and other mRNAs. *Plant Cell* 24 (3), 859–874. doi:10.1105/tpc.111.095380
- Song, L., Jiao, Y., Song, H., Shao, Y., Zhang, D., Ding, C., et al. (2023). NbMLP43 ubiquitination and proteasomal degradation via the light responsive factor NbBBX24 to promote viral infection. *Cells* 12 (4), 590. doi:10.3390/cells12040590
- Sun, H., Shen, L., Qin, Y., Liu, X., Hao, K., Li, Y., et al. (2018). Clc-nt1 affects potato virus Y infection via regulation of endoplasmic reticulum luminal pH. *New Phytol.* 220 (2), 539–552. doi:10.1111/nph.15310
- Sun, J., Wang, S., Li, C., Ren, Y., and Wang, J. (2014). Novel expression profiles of microRNAs suggest that specific miRNAs regulate gene expression for the sexual maturation of female *Schistosoma japonicum* after pairing. *Parasites Vectors* 7, 177. doi:10.1186/1756-3305-7-177
- Umbach, J. L., and Cullen, B. R. (2009). The role of RNAi and microRNAs in animal virus replication and antiviral immunity. *Genes Dev.* 23 (10), 1151–1164. doi:10.1101/gad.1793309
- Verchot, J. (2012). Cellular chaperones and folding enzymes are vital contributors to membrane bound replication and movement complexes during plant RNA virus infection. *Front. Plant Sci.* 3, 275. doi:10.3389/fpls.2012.00275
- Viswanath, K. K., Kuo, S. Y., Tu, C. W., Hsu, Y. H., Huang, Y. W., and Hu, C. C. (2023). The role of plant transcription factors in the fight against plant viruses. *Int. J. Mol. Sci.* 24 (9), 8433. doi:10.3390/ijms24098433
- Voinnet, O. (2009). Origin, biogenesis, and activity of plant microRNAs. *Cell* 136 (4), 669–687. doi:10.1016/j.cell.2009.01.046
- Whitworth, J. L., Nolte, P., McIntosh, C., and Davidson, R. (2006). Effect of potato virus Y on yield of three potato cultivars grown under different nitrogen levels. *PLANT Dis.* 90 (1), 73–76. doi:10.1094/PD-90-0073
- Yang, L., Mu, X., Liu, C., Cai, J., Shi, K., Zhu, W., et al. (2015). Overexpression of potato mir482e enhanced plant sensitivity to verticillium dahliae infection. *J. Integr. Plant Biol.* 57 (12), 1078–1088. doi:10.1111/jipb.12348
- Zhang, C., Ding, Z., Wu, K., Yang, L., Li, Y., Yang, Z., et al. (2016). Suppression of jasmonic acid-mediated defense by viral-inducible microRNA319 facilitates virus infection in rice. *Mol. Plant.* 9 (9), 1302–1314. doi:10.1016/j.molp.2016.06.014
- Zhang, N., Zhang, D., Chen, S. L., Gong, B. Q., Guo, Y., Xu, L., et al. (2018). Engineering artificial microRNAs for multiplex gene silencing and simplified transgenic screen. *Plant Physiol.* 178 (3), 989–1001. doi:10.1104/pp.18.00828
- Zhao, J. H., and Guo, H. (2022). RNA silencing: from discovery and elucidation to application and perspectives. *J. Integr. Plant Biol.* 64 (2), 476–498. doi:10.1111/jipb.13213



OPEN ACCESS

EDITED BY

Amaranatha Reddy Vennapusa,
Delaware State University, United States

REVIEWED BY

Suhas Gorakh Karkute,
Indian Council of Agricultural Research (ICAR),
India
Priyanka Siwach,
Chaudhary Devi Lal University, India
Rahul Nitnavare,
Rothamsted Research, United Kingdom
Harsha Vardhan Rayudu Jamedar,
Indian Institute of Rice Research (ICAR), India
Vishnutej Ellur,
Syngenta, United States

*CORRESPONDENCE

Muhammad Kashif Riaz Khan,
✉ mkrkhan@gmail.com
Yongming Liu,
✉ liuluckforever@gmail.com

[†]These authors have contributed equally to this work

RECEIVED 03 October 2023

ACCEPTED 01 February 2024

PUBLISHED 19 February 2024

CITATION

Nadeem S, Riaz Ahmed S, Luqman T, Tan DKY, Maryum Z, Akhtar KP, Muhy Ud Din Khan S, Tariq MS, Muhammad N, Khan MKR and Liu Y (2024), A comprehensive review on *Gossypium hirsutum* resistance against cotton leaf curl virus.
Front. Genet. 15:1306469.
doi: 10.3389/fgene.2024.1306469

COPYRIGHT

© 2024 Nadeem, Riaz Ahmed, Luqman, Tan, Maryum, Akhtar, Muhy Ud Din Khan, Tariq, Muhammad, Khan and Liu. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). 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.

A comprehensive review on *Gossypium hirsutum* resistance against cotton leaf curl virus

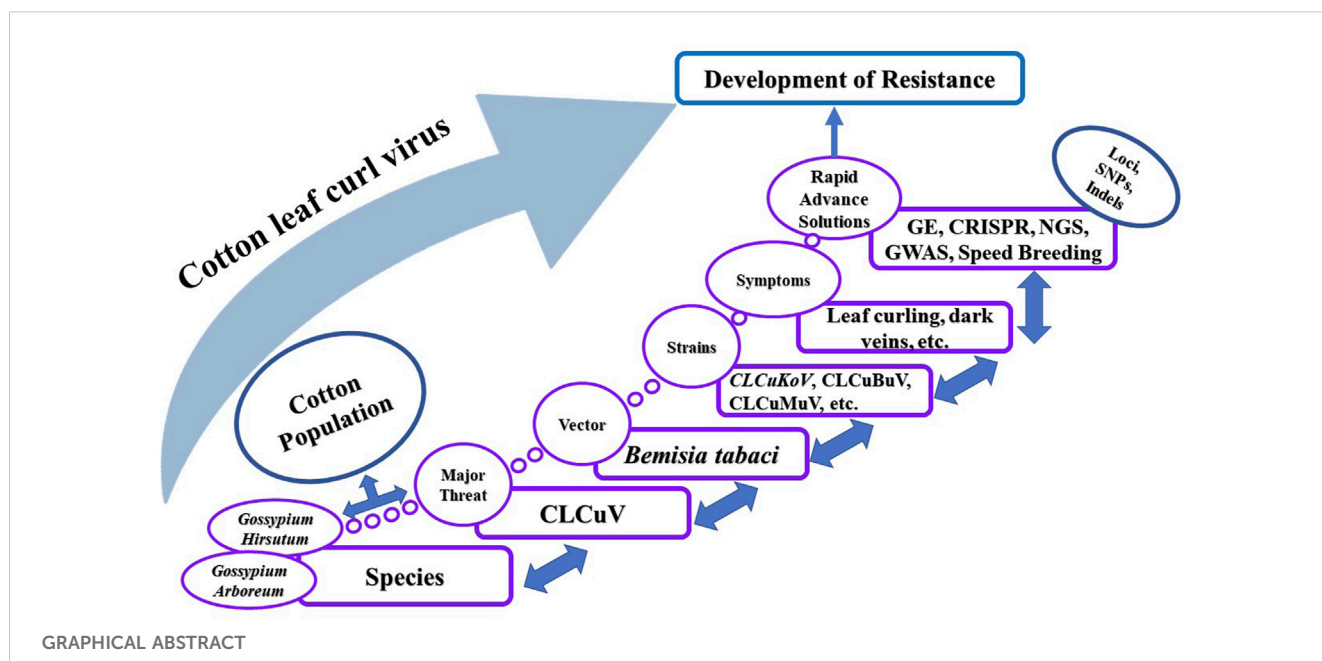
Sahar Nadeem^{1†}, Syed Riaz Ahmed^{1,2†}, Tahira Luqman¹, Daniel K. Y. Tan³, Zahra Maryum¹, Khalid Pervaiz Akhtar¹, Sana Muhy Ud Din Khan¹, Muhammad Sayyam Tariq¹, Nazar Muhammad⁴, Muhammad Kashif Riaz Khan^{1,5*} and Yongming Liu^{6*}

¹Nuclear Institute for Agriculture and Biology College, Pakistan Institute of Engineering and Applied Sciences (NIAB-C, PIEAS), Faisalabad, Pakistan, ²Pakistan Agriculture Research Council (PARC), Horticulture Research Institute Khuzdar Baghbana, Khuzdar, Pakistan, ³School of Life and Environmental Sciences, Plant Breeding Institute, Sydney Institute of Agriculture, Faculty of Science, The University of Sydney, Sydney, NSW, Australia, ⁴Agriculture and Cooperative Department, Quetta, Pakistan, ⁵Plant Breeding and Genetics Division, Cotton Group, Nuclear Institute for Agriculture and Biology, Faisalabad, Pakistan, ⁶National Nanfan Research Institute (Sanya), Chinese Academy of Agricultural Sciences, Sanya, China

Cotton (*Gossypium hirsutum* L.) is a significant fiber crop. Being a major contributor to the textile industry requires continuous care and attention. Cotton is subjected to various biotic and abiotic constraints. Among these, biotic factors including cotton leaf curl virus (CLCuV) are dominant. CLCuV is a notorious disease of cotton and is acquired, carried, and transmitted by the whitefly (*Bemisia tabaci*). A cotton plant affected with CLCuV may show a wide range of symptoms such as yellowing of leaves, thickening of veins, upward or downward curling, formation of enations, and stunted growth. Though there are many efforts to protect the crop from CLCuV, long-term results are not yet obtained as CLCuV strains are capable of mutating and overcoming plant resistance. However, systemic-induced resistance using a gene-based approach remained effective until new virulent strains of CLCuV (like Cotton Leaf Curl Burewala Virus and others) came into existence. Disease control by biological means and the development of CLCuV-resistant cotton varieties are in progress. In this review, we first discussed in detail the evolution of cotton and CLCuV strains, the transmission mechanism of CLCuV, the genetic architecture of CLCuV vectors, and the use of pathogen and nonpathogen-based approaches to control CLCuV. Next, we delineate the uses of cutting-edge technologies like genome editing (with a special focus on CRISPR-Cas), next-generation technologies, and their application in cotton genomics and speed breeding to develop CLCuV resistant cotton germplasm in a short time. Finally, we delve into the current obstacles related to cotton genome editing and explore forthcoming pathways for enhancing precision in genome editing through the utilization of advanced genome editing technologies. These endeavors aim to enhance cotton's resilience against CLCuV.

KEYWORDS

Cotton (*G. hirsutum* L.), CLCuV, genome editing, next-generation technologies, speed breeding



1 Introduction

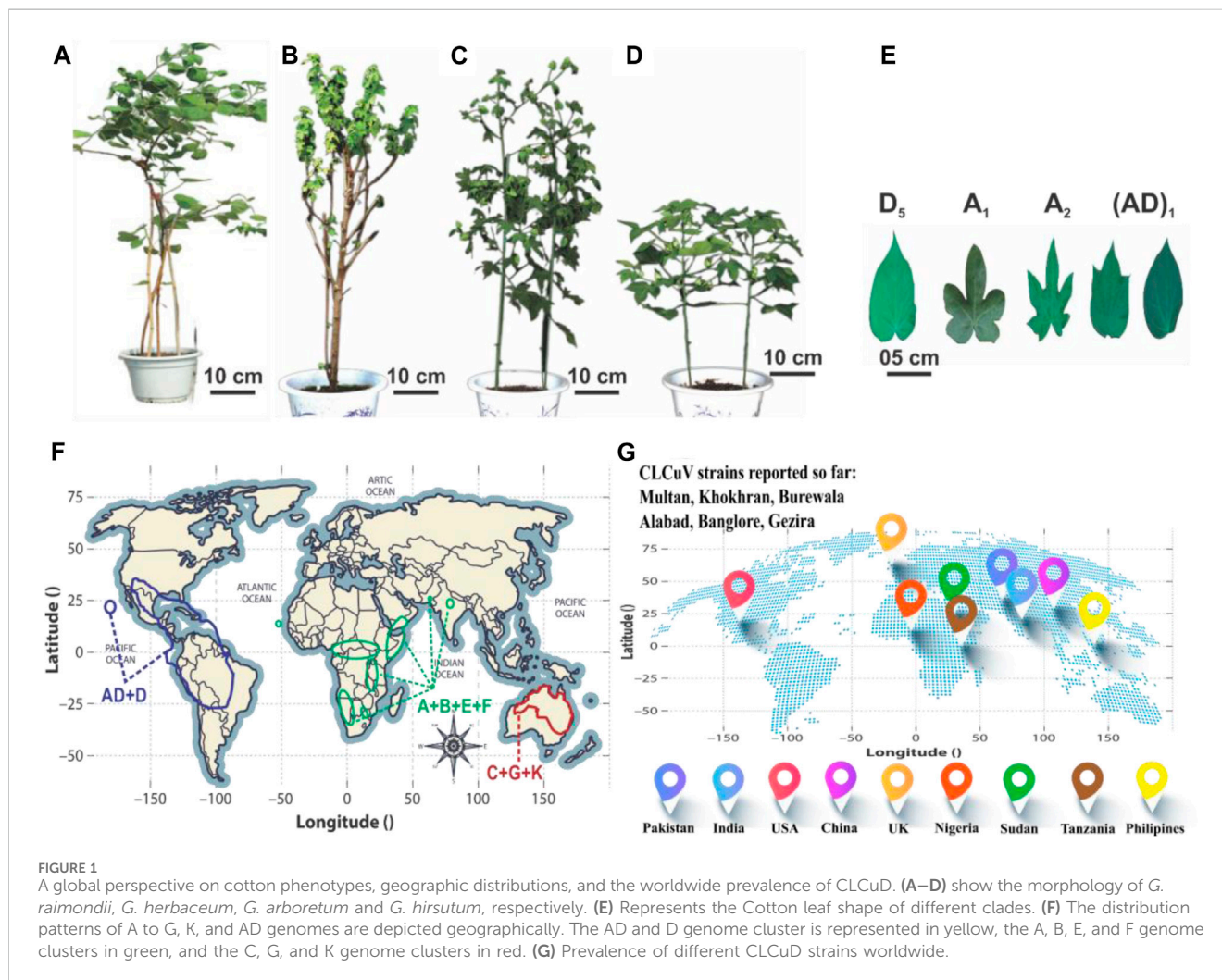
Cotton (*Gossypium hirsutum* L.) stands as a high-quality fiber-producing plant that contributes greatly to the world's textile industry by generating an annual economic income of 600 billion US dollars (Ashraf et al., 2018). It serves as a model system for investigating plant polyploidization, cell wall biogenesis, and cell elongation (Lu et al., 2018). The genus *Gossypium* encompasses seven tetraploid species ($2n = 4x = 52$) and 45 diploid species ($2n = 2x = 26$), showcasing remarkable morphological diversity (Lu et al., 2018). This variation includes a spectrum of plant structures, spanning from untamed perennial trees and shrubs to cultivated annual herbaceous plants, accompanied by unique fiber traits and diverse leaf shapes. Cotton has captivated the interest of agricultural researchers, evolutionary biologists, and taxonomists for its remarkable genomic diversity and widespread dispersion. This diversity has given rise to the evolution of eight distinct diploid cotton groups, designated as A-, B-, C-, D-, E-, F-, G-, and K-genomes, along with an AD-genome clade. The classification of the genus *Gossypium* into three primary lineages delineated mainly by geographical and morphological factors, including the Australian clade (C-, G-, and K-genomes), the African-Asian clade (A-, B-, E-, and F-genomes), and the New World clade (D- and AD-genomes) (Figure 1) (Ahmed, 1999; Huang et al., 2021). Among these, *G. hirsutum* [(AD)₁], commonly called upland cotton, currently dominates the global cotton industry by contributing approximately 95% of the natural lint fiber essential for textiles.

The textile industries rely on good quality cotton lint fiber to fulfill market demands. Each clade has a separate fiber lint quality and yield. Therefore, the fiber lint is of great importance in cotton domestication and quality among cotton breeders. The progression of fiber development can be categorized into four successive phases: initiation, elongation, secondary cell wall (SCW) biosynthesis, and maturation, categorized based on the days post-anthesis (DPA)

(Prasad et al., 2022). Cotton fibers can be additionally classified into two forms: adherent fuzz fibers, which commence development around 5 to 10 DPA and reach a final length of about 5 mm, and spinnable lint fibers, originating before flowering and growing to a final length of roughly 3 cm.

Pakistan is the fourth-largest producer of cotton following India, China, and the United States of America (Shuli et al., 2018). Unfortunately, cotton production is continuously being challenged by various biotic and abiotic stresses, among which CLCuV is most prevalent. CLCuV has emerged as a prominent threat to cotton, causing major losses of up to 80%–87% in Pakistan and North India (Mansoor et al., 2003; Varma and Malathi, 2003; Sattar et al., 2013; Monga and Sain, 2021; Chauhan et al., 2023). The importance of cotton in the textile industry is undeniable, therefore, requires continuous attention to its improvement. The virus is a major cause of low productivity in cotton and hence, efforts are being made to induce resistance in cotton crop against the disease (Ali et al., 2015). The story of the emergence of CLCuV as a threat is not new. For the first time, it was reported in Nigeria in 1912. Later on, it was documented in Sudan in 1924, followed by occurrences in Tanzania in 1926, and later in the Philippines in 1959. In Pakistan, CLCuV was discovered in 1967 on scattered cotton plants of the *G. hirsutum* species in the Multan district (Hussain and Ali, 1975). However, the disease was not considered a serious threat until the 1990s, when it appeared as an epidemic and disastrously brought a monetary loss of \$5 US billion, throughout the world (Bridson and Markham, 1994; Leke et al., 2015).

CLCuV is acquired, carried, and transmitted via whitefly (*Bemisia tabaci*) (Nogia et al., 2014). Once the vector acquires the virus, it is retained by the carrier whitefly throughout its life. Plants infected with CLCuV show a broad range of symptoms ranging from stunted growth, yellowing of leaves, and appearance of dark green thickened veins on leaves followed by upward or downward curling (Heigwer et al., 2016). In case of high



disease severity, curled leaves may even develop small outgrowths (also called enations) on their lower surface. Moreover, affected leaves become brittle and leaf size is substantially reduced (Mansoor et al., 1993; Harrison et al., 1997; Qazi et al., 2007). At the seedling stage, CLCuV affects the cotton yield by reducing the cotton boll's number and size, flowering time, and posing undesirable effects on seed and fiber quality, and final yield (Ali et al., 2013; Ji et al., 2015; Heigwer et al., 2016). It is challenging to fight CLCuV because of the prevalence of active viral strains and greater recombination rates of CLCuV complex. A case study by Qadir et al. (2019), revealed that the recombination potential of CLCuV is significantly higher. A greater recombination rate was observed in all genes (*IR*, *V1*, *V2*, *C1*, *C4*, and *C5*) associated with cotton leaf curl Kokhran Virus (CLCuKoV). Furthermore, CLCuKoV also donated in the *C2*, *C3* regions of cotton leaf curl Multan virus (CLCuMuV). Entirely, these observations clearly show the uniqueness of Indian CLCuMuV isolates representing the contribution of CLCuKoV in all the genes (Qadir et al., 2019). Higher recombination rate of CLCuV leads to multiple strain infection which makes viral control difficult. The presence of various alternative hosts and favorable environmental conditions has further complicated the disease control. Many efforts were made during the 1990s to produce CLCuV resistant varieties, yet the outbreak of another CLCuV

strain, namely, Cotton Leaf Curl Burewala (CLCuBuV) broke the resistance and the varieties fell prey to it (Sarwar et al., 2022).

Note: Cotton in Asia, experienced two significant epidemics within this time frame: the 'Multan epidemic,' spanning from 1988 to 1999, followed by a period of relative calm. However, in 2002, the 'Burewala epidemic' emerged in the cotton fields of the Indo-Pak subcontinent and persisted until 2013–2014.

The availability of various alternative hosts like hibiscus, okra, etc., and different farming practices could provide a reservoir of novel sources of new virulent CLCuV strains (Ali et al., 2022). Systemic-induced resistance using a gene-based approach remained effective until a new virulent strain of CLCuV came into existence named Cotton Leaf Curl Burewala Virus (CLCuBV). The damaging effects of the virus could only be overcome by proper management strategies. Many efforts have been used to protect the cotton crop from CLCuD. Disease control by biological means and development of CLCuV resistant cotton varieties are in progress. Use of pathogen and non-pathogenbased approaches and advanced molecular practices can be an effective way to fight viral causative agents (Hasan et al., 2019). We present an insightful overview encompassing the comprehension of the cotton genome, unraveling the genetic architecture of CLCuV, understanding cotton's defense mechanisms against CLCuD, and exploring the

applications of genome editing techniques for disease management, with a focus on cotton. Furthermore, we delineate the broader applications of other innovative technologies such as Next-Generation technologies, functional genomics, and speed breeding in cotton biological research and enhancement. Lastly, we discuss the ongoing challenges related to cotton genome editing and deliberate on the future opportunities involving various strategies and advancements in genome editing tools to foster the sustainability of cotton production.

2 Navigating CLCuV vectors: unveiling transmission mechanisms

Over the past three decades, insects from the genus *Bemisia* have become the major devastating pests of agriculture and horticulture (Ribeiro et al., 2003). The most significant and widespread among these are the whiteflies that are associated with *Geminiviruses* or more specifically to *Begomoviruses*. Their transmission may be directly associated with the silver leaf type of the whitefly. *Bemisia tabaci* is an indiscriminate feeder, facilitating rapid and proficient spread of CLCuV from affected plant hosts to neighbouring crops (Brown and Czosnek, 2002). Among all damaging pests of cotton, whitefly is responsible for about 50% yield loss and is a major restraint of boll formation (Mansoor et al., 2003). Intensive escalation in the whitefly population is because of the immense propagation of diseases by viruses through whiteflies.

Begomoviruses exist in either mono-partite or bi-partite forms. Bi-partite *Begomoviruses* contain both DNA A and DNA B, associated with New World pathogens. On the other hand, mono-partite *Begomoviruses* only contain a circular single-strand DNA-A and two satellite molecules viz., α and β satellite, classified as an essential component for disease severity in Old World pathogens (Hasan et al., 2019). An example of a mono-partite particle is CLCuV (Liu et al., 1998; Mansoor et al., 1999). PCR amplification of CLCuV-infected cotton indicated the presence of CLCuV (Fiallo-Olivé and Navas-Castillo, 2023). CLCuV possesses a circular single-stranded deoxyribonucleic acid (ssDNA) molecule, tightly encapsulated inside a geminate particle. Previous research has shown a wide range of hosts for mono-partite and bi-partite *Begomoviruses* (Yazdani-Khameneh et al., 2016). As the most destructive pathogens, CLCuVs are widely spread in Central and South-East Asia. The viruses carry distinct recombinations to break down transgenic resistance in cotton cultivars (Devendran et al., 2022).

2.1 Geminiviruses

Geminivirus is a group of pathogenic viruses transmitted by insects. The family *Geminiviridae* is responsible for causing several diseases in plants around the globe (Arif et al., 2022). *Geminiviruses* contribute to the largest group of plant viruses. Zerbin and colleagues (2017), categorized *Geminiviruses* into nine genera, utilizing criteria such as genome arrangement, transmission vectors, and host plants. This classification comprises *Turncurtovirus*, *Topocuvirus*, *Mastrevirus*, *Grablovirus*, *Eragovirus*, *Curtovirus*, *Capulavirus*, *Begomovirus*, and *Becurtovirus* along with a subset of species that remain unclassified (Zerbin et al., 2017). Recently, this family has been increased from nine recognized genera to fourteen genera (Materatski

et al., 2021). *Begomoviruses* is the largest group of *Geminiviruses* with around 520 accepted viral species ((Uniyal et al., 2019; Ouattara et al., 2022; Fiallo-Olivé and Navas-Castillo, 2023).

CLCuV *Geminiviruses* are responsible for the transmission of CLCuV through whiteflies. These viruses are taken up by whiteflies when they feed on infected plants and are then transmitted to healthy plants during subsequent feeding. This transmission mechanism makes *Geminiviruses* key contributors to the spread and persistence of CLCuV in cotton crops, leading to significant agricultural and economic impacts (Jain et al., 2023). Significant crop losses (such as high plains viral disease of wheat and corn cause by wheat mosaic virus and maize red stipe virus, sorghum mosaic virus and sugarcane mosaic virus cause by potyvirus, and rice stripe virus cause and transmitted by *Laodelphax striatellus*) have been reported due to various disease infestations leading to viral diseases spread worldwide (Lu et al., 2021; Tatineni and Hein, 2021; Xu, et al., 2021). Major reasons that resulted in disease epidemics include the recombination of many different *Geminiviruses* (i.e., the exchange of genetic materials among at least two *Geminiviruses* to become more virulent) that co-infect the same crop, expansion, and development of agriculture in newer areas, or transport of infectious plant materials to other regions. Additionally, plant co-infections can be spread by vectors that pick up the pathogens from plants infected with multiple diseases, either from different plants one after another, or by different vectors each carrying a different disease. Moreover, the migration of carrier insect vectors from one location to another also brings epidemic explosion (Gray and Banerjee, 1999).

2.2 Begomoviruses

The genus *Begomovirus* is a serious threat not only to cotton but also to other crops such as tomato and cassava (Donnelly and Gilligan, 2023; Fortes et al., 2023). *Begomovirus* behaves the same way as *Geminiviruses* in transmitting the virus. The complex interaction between *Begomoviruses*, whiteflies, and cotton plants highlights the significance of understanding and managing these factors to mitigate the impact of CLCuV on cotton crops. Bi-partite begomoviruses are comprised of DNA-A and DNA-B, each approximately 2,600 nucleotides long (Fondong, 2013). The mono-partite particles contain single DNA-A molecule of about 2,800 nucleotides (Fauquet and Stanley, 2003). *Betasatellites* encompass circular, single-stranded DNA molecules comprising around 1,350 nucleotides, responsible for disease inducing symptoms (Zhou et al., 2003; Saunders et al., 2004). *Alphasatellites* are also circular single-stranded DNA made up of approximately 1,375 nucleotides and associated with a rolling-circle replication (RCR) initiator protein.

3 Decoding CLCuV genome: unveiling the genetic architecture

3.1 DNA A

CLCuV is an example of a mono-partite particle that contains single-stranded DNA (Briddon et al., 2000). DNA-A of mono-partite particles encodes for viral DNA replication, insect transmission and control gene expression (Noueiry et al., 1994). DNA-A has six open

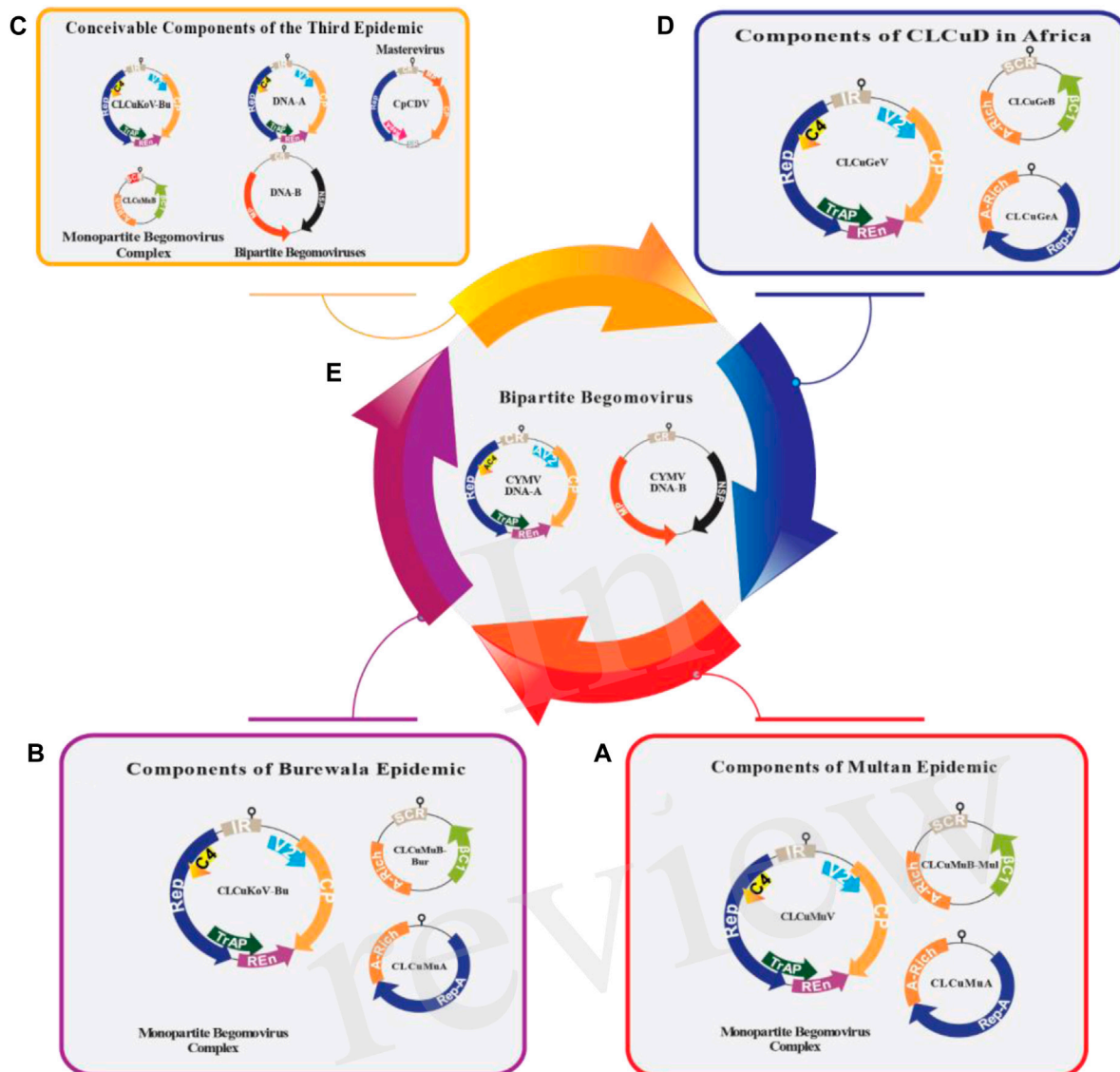


FIGURE 2

The genomic arrangement of all the contributing viruses responsible for CLCuD in Asia (A–C) and Africa (D,E). The prominent *Begomovirus*, along with DNA satellites associated with the suggested third epidemic (C), Burewala epidemic (B) and Multan epidemic (A) are displayed. In the period of the Multan epidemic, various *Begomoviruses* were simultaneously identified in cotton plants, including tomato leaf curl Bangalore virus (ToLCBaV), Papaya Leaf Curl Virus (PaLCuV), cotton leaf curl Kokhran virus (CLCuKoV) and cotton leaf curl Alabad virus (CLCuAIV). Nevertheless, the primary.

reading frames (ORFs) that encode various proteins (Yaqoob et al., 2020). In the virion sense, there are two ORFs, namely, AV2 (encodes the movement of a protein; also called precoat) and AV1 (encodes the coat protein), while in the complementary sense, four ORFs [AC4 (determines the expression of symptoms), AC3 (encodes replication enhancer protein), AC2 (encodes a transcription activator protein), and AC1 (encodes the initiation protein called Rep)] are found (Marwal et al., 2014). The genomes of mono-partite viruses and the DNA-A component of bi-partite *Begomoviruses* contain the instructions for the coat protein (CP) and V2 (also known as AV2) in the orientation corresponding to the virion sense (Figure 2). The coat protein is involved in virus proliferation, vector transmission, disease virulence and encapsidation (Singh et al., 2021). AV2 is associated with cell-to-cell movement and pathogenicity (Arif et al., 2022). The proteins linked to

replication (AC1/C1 or Rep protein), transcriptional activation (AC2/C2 or TrAP protein), replication enhancement (AC3/C3 or Ren protein), and the AC4/C4 protein are encoded in the orientation corresponding to the complementary sense (Fondong, 2013; Zhou, 2013). AC1 serves viral replication and gene expression whereas AC2 is responsible for transcription activators of rightward ORFs and suppressors of post-transcriptional gene silencing (PTGS) (Dilip, 2016; Sharma and Prasad, 2020). AC3 is involved in viral replication and symptom development and AC4 in gene silencing (Faiz and Abhinav, 2021; Roumagnac et al., 2021). DNA-B consists of two ORFs, BC1 in the complementary sense and BV1 in the virion sense. BV1 plays a role in nuclear trafficking and BC1 encodes movement protein MP, involved in inter- and intracellular movement (Diamos et al., 2019; Yadav et al., 2021).

3.2 DNA B

The genome component DNA-B (absent in CLCuV but present in bi-partite *Begomoviruses*) encode the movement protein (MP) in the complementary-sense orientation and nuclear shuttle protein (NSP) in the virion sense (Sahu et al., 2014). DNA-B has two open reading frames and encodes proteins associated with inter and intracellular movements (Mubarik et al., 2021). An intergenic intronic region exists between ORFs of the virion and complementary-sense regions that contain a hair-pin loop-like structure and cis-acting regulatory elements essential for gene expression. The loop-like structure contains a nona-nucleotide conserved sequence 'TAATATTAC' and small repetitive sequences 'iterons' that provide sequence-specific binding sites for Rep (C_1/AC_1) protein. Both iterons and hairpin loop form the origin of replication (*ori*) for initiation of viral DNA replication (Zhou, 2013).

3.3 DNA α -satellite and β -satellite

Alpha and *betasatellites* are assumed to function in encapsulation, insect mediated transmission and cellular movement inside the infective host plants (Gnanasekaran and Chakraborty, 2018). Additionally, *betasatellites* impact symptom development and boost the pathogenicity of helper viruses by elevating viral DNA levels in host plants and inhibiting the plant's antiviral defense mechanisms (Gnanasekaran et al., 2019). Satellite molecules share a conserved loop like structure with DNA-A. Genome size of these molecules are about 1.4 kb (Azeem et al., 2022). Both satellites contain conserved structures which share broad characteristics. Furthermore, the molecules comprise an "Ori" region separated by a non-nucleotide sequence "NANTATTAC" at the apex of the hairpin structure (Rosario et al., 2012). Though, *betasatellites* contain a nonanucleotide motif "TAATATTAC", observed in a large number of *Begomoviruses* whereas *alpha satellites* contain "TAGTATTAC" motif, seen in various members of *Nanoviridae* family. *Alphasatellite* is capable of replicating itself autonomously. Both *alpha* and *betasatellites* are rich in adenine residues which function in proficient encapsulation and aid in systemic movement of the virus. Both molecules encode a single protein. The α -satellite encodes a Replication protein (RP) associated with rolling circle replication inside the host cells. On the contrary, helper *Begomovirus* is required for *betasatellite* replication. The β -satellite also encodes $\beta C1$ protein which is multi-functional and significant for inducing pathogenicity (Zhou, 2013).

Begomovirus responsible was for Cotton Leaf Curl Multan Virus (CLCuMuV). In the case of the Burewala epidemic, a recombinant cotton leaf curl Multan *betasatellite*-Burewala strain (CLCuMuB^{Bur}) and a truncated transcriptional activator protein (TrAP) were observed, incorporating approximately 98 nucleotides from the SCR region acquired from Tomato leaf curl *betasatellite* (ToLCB) (B). The third epidemic in Asia presents several possible scenarios: i) it could involve Cotton leaf curl Kokhran virus-Burewala (CLCuKoV-Bu), which retains an intact TrAP, along with Cotton leaf curl Multan *betasatellite* (CLCuMuB), containing a ~24 nt stretch from tomato leaf curl *betasatellite* (ToLCB), ii)

Alternatively, it might include tomato leaf curl New Delhi virus (ToLCNDV) (comprising both DNA-A and DNA-B) and tomato leaf curl virus (ToLCV) (with DNA-A), possibly in conjunction with ToLCNDV (DNA-B), iii) Another potential contributor could be Mastrevirus chickpea chlorotic dwarf virus (CpCDV) (C). Concurrently, the CLCuD situation in Africa comprises: (i) a mono-partite *Begomovirus* complex (D), a recently identified bipartite *Begomovirus*, cymbidium mosaic virus (CYMV), with DNA-A and DNA-B (E). The DNA-A of mono-partite *Begomoviruses* encodes two open reading frames (ORFs) in the virion-sense orientation, specifically the V2 protein and the coat protein. In the complementary-sense orientation, there are four ORFs: the replication-associated protein (Rep), TrAP, replication enhancer protein (REN), and C4 protein (with functions yet to be fully understood). Bi-partite *Begomovirus* DNA-A resembles the mono-partite genome but includes a second genomic component, DNA-B, which encodes two ORFs: the movement protein (MP) and the nuclear shuttle protein (NSP), in both complementary-sense orientations and virion sense. The *alphasatellite* encodes a single ORF (Rep-A) for autonomous replication, while the *betasatellite* encodes its single ORF, $\beta C1$, in the complementary-sense orientation. $\beta C1$ plays a role in assisting the helper *Begomovirus* in various functions.

4 Defense mechanisms of cotton against CLCuV

Besides viruses, several other organisms including insects such as cotton aphids, *Hemipteran* sucking insects, cotton bollworms, and *Lepidopteran* chewing caterpillars are commonly observed in cotton fields. Cotton bollworm predation triggers the activation of genes related to the gibberellic acid (GA) (*GhMPK11*, *SLR1*), ethylene (*GhWRKY70D13*, *GhERF91*), and jasmonic acid (JA) (*GhWRKY70D13*, *JAZ3*) pathways, while simultaneously suppressing genes associated with the salicylic acid pathway within cotton plants (Kumar et al., 2016; Wang et al., 2016; Xia et al., 2018; Xiong et al., 2020; Kundu et al., 2023; Nurimanguli et al., 2023). JA acts as a regulator in facilitating insect defense, working in conjunction with its corresponding target (SPL9) and miR156 (Arora and Chaudhary, 2021). The response to JA is swift in young plants but progressively slows as plants age, displaying an age-related pattern inversely linked to SPL9 group protein levels. Moreover, not all but some of the herbivorous insects have also developed intricate strategies, such as the discharge of effector molecules into the host and the synthesis of diverse detoxification enzymes, to overcome the resistance mechanisms of their host plants. The defense mechanisms of cotton against CLCuD are multifaceted and include both pathogen-derived and non-pathogen-derived strategies. The most significant defense mechanism is the presence of resistance (R) genes in plants (Table 1). These genes provide specific resistance against particular viruses by inducing cell death around infected plant cells, thereby preventing the movement of the virus. Cotton has two associated R genes, *R1CLCuDhir* and *R2CLCuDhi*, which confer resistance against CLCuD (Khan et al., 2015). Furthermore, the exclusive protein $\beta C1$, encoded by the satellite, has been demonstrated to hinder the ubiquitination process of the host

TABLE 1 List of genes associated with CLCUD and other biotic and abiotic factors, directly and indirectly, involved in cotton infections.

Sr. No	Gene ID/ Accession no.	Gene symbol	Gene description	Chromosomal position/locus tag	Gene function	References
1	30144432	AV1	Coat protein	BMU10_gp3	Virus proliferation, vector transmission, disease virulence and encapsidation	Singh et al. (2021)
2	30144433	AV2	Pre-coat protein	BMU10_gp2	Involved in Pathogenicity and cell to cell movement	Sodha et al. (2022)
3	30144436	AC1	Replication associated protein	BMU10_gp6	Viral replication and gene expression	Sharma and Prasad. (2020)
4	30144435	AC2	Transcription activator protein	BMU10_gp5	Responsible for transcription activators of rightward ORFs. Suppressor of post-transcriptional gene silencing (PTGS)	Dilip (2016)
5	30144434	AC3	Replication enhancer protein	BMU10_gp4	Viral replication and symptom development	Roumagnac et al. (2021)
6	30144437	AC4	Regulatory protein	BMU10_gp7	Gene silencing	Faiz and Abhinav. (2021)
7	80557347	βC1	BetaC1 protein	QKT41_gp1	Symptom determinant, interact with other helper viruses, affect number of plant species including cotton	Iqbal et al. (2012)
8	988208	ac5	AC5 protein	ToLCNDVsAgp4	Viral replication	Amudha et al. (2010)
9	107958459	WRKY4 0	Probable WRKY transcription factor 40	Chromosome A05/ ERO13_A05G119 500v2	Regulate cotton defense for controlling whitefly infestation	Li et al. (2016)
10	107924381	MPK	Mitogen-activated protein kinase kinase 3	Chromosome A11	Silencing of MPK leads to enhanced whitefly susceptibility	Li et al. (2016)
11	988213	BV1	Nuclear shuttle protein	ToLCNDVsBgp1	Role in nuclear trafficking	Diamos et al. (2019)
12	2767401	BC1	BC1 protein	TYLCKV_sBgp2	Encodes movement protein MP, involved in inter- and intracellular movement	Yadav et al. (2021)
13	—	<i>R1CLCuDhir</i>	Resistance gene	—	Provides resistance to cotton against CLCuD	Hagan et al. (2019)
14	—	<i>R2CLCuDhir</i>	Resistance gene	—	Provides resistance to cotton against CLCuD	Hagan et al. (2019)
15	—	<i>SCLCuDhir</i>	suppressor gene	—	Suppressor of resistance to CLCuD	Hagan et al. (2019)
16	107934102	LOC107 934102	Probable zinc metallopeptidase EGY3, chloroplastic	ERO13_A01G009 420v2/ Chromosome A01	Chloroplast development. Copes with stress	Naqvi et al. (2017)
17	107928973	LOC107 928973	Probable zinc metallopeptidase EGY3, chloroplastic	ERO13_D01G009 000v2/ Chromosome D01	Development and stress response	(Naqvi et al. 2017)
18	105786625	LOC105 786625	Probable zinc metallopeptidase EGY3, chloroplastic	B456_002G01100 0/ chromosome 1	Probable membrane-associated metalloprotease that may be involved in chloroplast development.	Marino and Funk. (2012)
19	107909622	LOC107 909622	Chaperone protein dnaJ 16	ERO13_D02G166 200v2/ chromosome D02	Pathogen defense, vital in antiviral defense	Verchot (2012), Park and Seo (2015)
20	105802570	LOC105 802570	Transcription factor UNE10	B456_007G08830 0/ chromosome 11	Antiviral defense	Wang et al. (2015)
21	107911889	LOC107 911889	Transcription factor UNE10- like	ERO13_A11G078 200v2/ chromosome A11	Antiviral defense	Wang et al. (2015)
22	107904753	LOC107 904753	Protein REVEILLE 7	ERO13_D12G273 200v2/ chromosome D12	Plant growth contributes to response against pathogens	Wang et al. (2015), Zhong et al. (2016)

(Continued on following page)

TABLE 1 (Continued) List of genes associated with CLCUD and other biotic and abiotic factors, directly and indirectly, involved in cotton infections.

Sr. No	Gene ID/ Accession no.	Gene symbol	Gene description	Chromosomal position/locus tag	Gene function	References
23	107902644	LOC107 902644	Transcription factor RADIALIS	ERO13_D05G196 766v2/ Chromosome D05	Plant defense response	Katiyar et al. (2012)
24	107951700	LOC107 951700	Putative disease resistance protein RGA3	ERO13_D11G336 100v2/ chromosome D11	R-gene-mediated resistance and disease response	Chen et al. (2015)
25	107916180	LOC107 916180	Bidirectional sugar transporter SWEET17	ERO13_D10G171 600v2/ chromosome D10	Confer's tolerance against abiotic stress, pathogenesis-related protein	Chandran (2015)
26	105792917	LOC105 792917	Protein ECERIFERU M 16	B456_004G28270 0/ chromosome 8	Role in biotic and abiotic stresses	Bourdenx et al. (2011)
27	107928876	LOC107 928876	Aquaporin PIP2-7-like	ERO13_D01G001 900v2/ chromosome D01	Critical in plant immunity, biotic and abiotic stresses	Matsumoto et al. (2009), Meng et al. (2016), Tian et al. (2016)
28	107927933	LOC107 927933	Potassium channel AKT2/3	ERO13_A03G193 300v2/ chromosome A03	Plant development, stress responses, antiviral defense	(Wang and Wu 2010, Zhou et al. 2014)
29	107937358	LOC107 937358	Boron transporter 4	ERO13_A06G187 400v2/ chromosome A06	Involved in R-gene-mediated viral defense	Lim et al. (2014)
30	107949718	LOC107 949718	Vignain	ERO13_D03G102 700v2/ chromosome D03	Plant immunity, pathogenesis, and plant defense	Figueiredo et al. (2014)
31	SAMS	GhSAMS	S-adenosylmethionine synthase	Chromosome AD1	Drought tolerance	Sun et al. (2022)
32	TR104989_c0_g1	LSM14-like protein B	Zf-C2H2 transcription factors	—	CLCuV resistance	Farooq et al. (2022)
33	AT5G11270.1	Gh_D08G2462	A homeodomain transcription factor	Chromosome _D8	Potential candidate gene for drought tolerance	Guo et al. (2022)
34	AT2G29560.1	Gh_A13G1963	A putative Phosphoenolpyruvate enolase	Chromosome A13	Potential candidate gene for drought tolerance	(Guo et al. 2022)
35	—	GhCFIm25	Pre-mRNA cleavage factor Im 25 kDa subunit	—	A potential candidate gene for Aphis gossypii tolerance	Yang et al. (2023)
36	—	GhGUX5	Involved in several physiological processes		A candidate gene associated with resistance to <i>Verticillium</i> wilt	Zhang et al. (2023)

plant (Gnanasekaran et al., 2019). It achieves this by interacting with the SKP1/CUL1/F-box complex, disrupting plant hormone signaling (Zhang Y. et al., 2019). This disruption ultimately leads to the buildup of the virus and the clustering of symptoms in the host plants. Moreover, βC1 also disrupts the plant's autophagy pathway by obstructing the normal function of autophagy-related protein 3 and cytosolic glyceraldehyde-3-phosphate dehydrogenase within the host cells (Ismayil et al., 2020).

4.1 Pathogen derived resistance strategies

RNA interference (RNAi) is another potent defense mechanism in plants. This mechanism involves the degradation of viral mRNA, effectively silencing the genes responsible for viral replication. For example, after the initial cloning of the first effector glucose oxidase from *Helicoverpa zea* (Musser et al., 2002), many other effectors like HARP1 (Chen et al., 2019), and cytochrome P450 monooxygenases (TAO et al., 2012), were discovered. In addition to phytochemicals, phytohormones (SA and JA) also have the capability to activate insect P450 genes when insects invade (Li et al., 2002). Both

alterations in the structure of P450 proteins and increases in the expression of P450 genes have been linked to the ability of insects to tolerate gossypol and insecticides such as fenvalerate and deltamethrin (Joußen et al., 2012). Upon introducing a double-stranded RNA (dsRNA) construct that targeted the gossypol-inducible P450 gene *CYP6AE14* (referred to as *dsCYP6AE14*) into *Arabidopsis* plants, there was a significant reduction in the expression of *CYP6AE14*, resulting in a simultaneous decrease in the weight gain of larvae (Mao et al., 2007). Similarly, employing dsRNA constructs to target genes like *dsFAR* (fatty acyl-CoA reductase), *dsJHBP* (JH-binding protein), *dsJHAMT* (juvenile hormone acid methyltransferase) and *dsNDUFV2* (encoding a mitochondrial complex I subunit) in the sap-sucking pest *Adelphocoris suturalis* of the *Miridae* family and *Helicoverpa armigera* yielded positive outcomes when transformed into cotton plants for pest control (Huang et al., 2021). Additionally, plants produce natural defense chemicals like nitric oxide (NO), salicylic acid (C₇H₆O₃), and reactive oxygen species (ROS) to combat pathogens (Ibrahim et al., 2023). Short interfering RNAs (siRNAs) also play a role in plant defense against double-stranded RNA, although some viruses have developed proteins to counteract

this response. In the event of injury, plants restrict viral movement through plasmodesmata. These findings now urge cotton breeders to use RNAi extensively in controlling CLCuD.

The primary approach to controlling CLCuD involves managing the whitefly population through insecticides. However, this method is not entirely reliable, as delayed application could lead to transmission of the infectious strain. Moreover, insecticide use has adverse environmental effects and can be toxic to various organisms. Therefore, conventional methods of protecting cotton crops from insects are not preferable. Biotechnological and molecular breeding approaches present the most promising option for inducing resistance in cotton against viruses (Saakre et al., 2023). Pathogen-derived resistance can be further categorized into two types: resistance with protein expression and resistance without protein expression. Proteins play a crucial role in plant defense against biotic stresses, with research indicating that proteins act as the primary defense against CLCuD in cotton (Chauhan et al., 2023). Viral replication, regulated by Rep proteins, can be disrupted to inhibit the synthesis of viral components. However, while efforts have been made to develop transgenic cotton varieties against CLCuD, these varieties are still undergoing testing. The second approach, without protein expression, involves inducing resistance through RNAi. RNAi is a sequence-specific defense mechanism that leads to the degradation of viral mRNA transcripts. In plants, RNAi can be triggered by introducing a hairpin loop structure with a sequence homologous to the virus, thus stimulating the silencing signal (Karthik et al., 2023). This approach could also aid in developing insect-resistant genes against whitefly-transmitted viruses. RNAi-based resistance has been successful against Bean Golden Mosaic Virus in beans in Brazil (Souza et al., 2018). Another approach involves coat protein (CP) expression from other viruses, such as the Tobacco Mosaic Virus (TMV), to induce resistance. For example; Rasool et al. (2021), in their study, used synthetic genes to explore pathogen-derived resistance (PDR) for managing CLCuKoV-Bu and its associated betasatellite linked to CLCuD. Synthetic coat protein (CP_{syn}) and replication-associated protein (Rep_{syn}) genes from CLCuKoV-Bu were designed and inserted into *Nicotiana benthamiana* plants via *Agrobacterium*-mediated transformation. Transgenic plants carrying CP_{syn} and Rep_{syn} genes showed milder symptoms and lower virus concentrations when exposed to CLCuKoV-Bu or CLCuKoV-Bu-CLCuMuB inoculums. These findings suggest that CP_{syn} and Rep_{syn} genes have the potential for conferring resistance against CLCuKoV-Bu and related betasatellites and the study indicates the development of a PDR mechanism for controlling CLCuD. However, the exact mechanisms behind these strategies remain only partially understood (Rasool et al., 2021).

4.2 Non-pathogen derived resistance strategies

Non-pathogen-derived resistance strategies have been explored in cotton against CLCuD. The Need for non-pathogen-derived resistance strategies emerged when the applications of biotechnological approaches were found successful in other crops like wheat (Jastrzębska et al., 2020), rice (Montesinos et al., 2017), and maize (Gould and Northcote, 1986), but with a low success ratio

in cotton (Khan et al., 2015). These strategies include enhancing natural defenses like trichomes (which are hair-like structures, developed at the aerial plant surfaces involved in plant defense) and increasing the concentration of waxes and inorganic salts, which can protect cotton plants against whiteflies, CLCuD, and other pests. For example; Suthar et al., 2021, studied the tolerance response of interspecific derivatives (resulting from a cross between *Gossypium hirsutum* L. × *G. armourianum* Kearney) against whitefly. Their results revealed that the concentration of trichome at early developing stages was high and developed strong tolerance against whitefly. However, with age progression, a significant decline was observed in trichome concentration in mature leaves (Suthar et al., 2021). Majid et al. (2020) conducted an experiment to study the effect of epicuticular wax against CLCuD in different varieties of cotton crop. Their results revealed a strong significant negative correlation between leaf epicuticular wax and CLCuD infection, indicating that the higher the concentration of wax the lower will be the CLCuD infection (Majid et al., 2020).

5 Revolutionizing genetics: genome editing and beyond

5.1 CRISPR-Cas mediated genome editing for CLCuD

Genome editing (GE) technologies have brought about a transformative shift in the realm of plant research and hold immense promise for enhancing crop attributes. The Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)-CRISPR-associated protein (CRISPR-Cas) system, renowned for its adaptability, simplicity, and cost-effectiveness in achieving precise modifications to DNA sequences such as gene knockout or knockdown, allele/gene *in vivo* replacement, and single base substitution, has emerged as the dominant force in GE over recent years (Khan et al., 2022; Atiq et al., 2023). Upon creating double-strand DNA breaks (DSBs) by CRISPR-Cas, subsequent repair occurs via either the error-prone non-homologous end joining (NHEJ) pathway or the more precise homology-directed repair (HDR), or a combination of both pathways (Liu et al., 2022). NHEJ primarily facilitates DSB repair and typically induces random indels (also called insertions and deletions) at the rejoining point of the chromosomes. This propensity for indels has led to the predominant use of NHEJ in current plant GE endeavors for inducing gene knockouts and mutations. In the context of precise genome editing, HDR can be invoked through the presence of a DNA or RNA donor repair template (DRT) containing homologous sequences encircling the DSB. This permits more precise gene insertions or replacements (Tang et al., 2023). Alternatively, two other strategies also exist that still offer precise GE even without DRTs and DSBs. The first is called base editing (BE), a powerful technique facilitating gene replacement through HDR, enabling the precise substitution of a single base with a designated single nucleotide polymorphism (SNP) (Sony et al., 2023). This technique facilitates the precise alteration of plant genomes in a programmed manner. The second strategy, known as prime editing (PE), greatly extends the scope of precise genome editing by allowing all types of minor indels and base substitutions (Karlson et al., 2021).

Near in the future, the widespread adoption of various CRISPR toolsets will significantly hasten progress in both crop genetic improvement and biological research, particularly for those plant species with available reference genome sequences.

Until now, most of the research on genome editing in crop plants, such as cotton, has relied on NHEJ to create targeted loss-of-function mutations at particular gene locations. The applications of CRISPR-Cas9 have been widely employed to enhance both crop yield and quality such as resilience against abiotic and biotic factors including CLCuD. For instance; Binyameen et al. (2021), used multiplexed CRISPR-based GE system to develop resistance against CLCuD in cotton. They obtained the targeted CLCuD sequences from the National Center for Biotechnology Information (NCBI) and designed three guided RNA (gRNA) molecules to target three different sites in the genome. They successfully obtained 60%–70% resistance against CLCuD at the seedling stage compared with control plants (Binyameen et al., 2021). The β C1 and Rep genes were responsible for replication of CLCuV within the host genome. Using a newly developed multiplexed GE approach was more efficient and effective to knock down or knock-out these genes to develop resistance against CLCuD (Khan et al., 2020). Improving cotton quality and meeting the needs of the cotton industry is an important breeding objective for cotton breeders. The aim is to develop a cotton variety that not only showcases improved quality and resistance against CLCuD but also achieves increased production without sacrificing yield.

5.2 Base editing and its applications against CLCuD

BE offers a mechanism to achieve exact substitutions of one base pair for another at a designated genomic position, without necessitating a DRT and without instigating DSBs (Basu et al., 2023). This technique operates independently of the HDR pathway. Recently existing base editors, encompassing adenine base editors (ABEs) and cytosine base editors (CBEs), have been engineered by fusing the inactive Cas9 (dCas9) or Cas9 nickase (D10A) (nCas9) with adenine deaminase or cytidine deaminase, respectively (Komor et al., 2016; Gaudelli et al., 2017). In this manner, single-point mutations such as the conversion of A*T to G*C or C*G to T*A in crop plants can be generated using ABEs or CBEs (Bharat et al., 2020; Ali et al., 2023; Khan et al., 2023). So far, base editing has been effectively used to target several important traits in various crops like maize, cotton, *Arabidopsis thaliana*, rice and others. For example, Qin et al. (2020), developed the novel *G. hirsutum*-Base Editor 3 (GhBE3) to induce single base mutation in the allotetraploid genome of *G. hirsutum*. To test the efficiency of this system, two genes (*GhPEBP* and *GhCLA*) were targeted at three different sites in the genome. The efficiency of this base editing system was found highly effective and ranged from 26.67% to 57.78% (Qin et al., 2020). In another study, Wang et al. (2022), created diverse ABE vectors utilizing modified adenosine deaminase (TadA) proteins linked with two Cas9 variants named nCas9 and dCas9. These vectors demonstrated proficient A to G editing, achieving efficiencies up to 64% at targeted locations in the complex allotetraploid cotton genome. The results revealed

GhABE7.10n as the most effective, particularly altering the A5 position (considering PAM as positions 21–23). RNA off-target and DNA assessments using *GhABE7.10d* and *GhABE7.10n* edited cotton plants revealed minimal RNA off-target mutations and no DNA off-target changes. Additionally, a novel base editor, *GhABE7.10dCpf1*, was introduced, recognizing a PAM rich in thymine (T), leading to targeted A-to-G substitutions and a single amino acid modification in the cotton phosphatidyl ethanolamine-binding protein (*GhPEBP*). This alteration resulted in a condensed cotton plant structure that is ideal for mechanized harvesting in contemporary cotton farming (Wang et al., 2022). Recently, Mubarik et al. (2021) demonstrated a successful CRISPR-Cas9 system application to combat CLCuD by targeting overlapping genes of prevalent CLCuVs with three guide RNAs (gRNAs). Utilizing a multiplex CRISPR-Cas9 construct targeting six CLCuV genes proved more effective in inhibiting virus proliferation compared to individual targeting. Targeting CLCuV at multiple sites simultaneously resulted in superior interference and inefficient recovery of altered virus molecules. When tested in cotton plants, the multiplex construct significantly reduced virus accumulation in leaves. These findings highlight the potential of CRISPR-Cas9 for engineering virus resistance in crops, especially against mixed virus infections, through multiplex genome editing.

5.3 Prime editing and its applications against CLCuD

Prime editing, also known as PE, is a revolutionary genome-editing approach that allows precise modifications to the genetic code. Unlike traditional methods that rely on donor DNA templates or DSBs, prime editing utilizes a fusion protein complex to introduce specific edits at desired gene loci in plant cells (Hillary and Caesar, 2022). Prime editing involves the conjugation of a reverse transcriptase Moloney murine leukemia virus reverse transcriptase (M-MLV-RT) with a catalytically impaired Cas9 (nCas9) to form a protein complex. A guide RNA for prime editing (pegRNA) is employed to designate the desired target location and encode the intended genetic information. Within the prime editing system, the protein complex attaches to the targeted DNA sequence and creates a nick in the non-target strand (Li et al., 2021). Prime editing has gone through several iterations, with each version aiming to improve product purity and editing efficiency. The initial version, PE1, was developed by fusing the C-terminus of nCas9 with M-MLV-RT. This fusion protein showed promise but lacked optimal efficacy. To improve PE efficacy, PE2 was developed by replacing the original M-MLV-RT with an engineered version containing six mutations: *L603W*, *T330P*, *W313F*, *T306K*, *D200N*, and *H9Y*. Building upon the success of PE2, PE3 was created by introducing another nicking sgRNA at various distances from the nicks induced by the pegRNA. This additional cleavage in the non-edited strand further increased the efficiency of the editing process (Anzalone et al., 2019).

PE shows significant potential for introducing precise genome modifications in cotton crop to enhance agronomic traits and develop resistance against CLCuD. In cotton, CLCuD tolerant plants were developed using prime editing (Cao et al., 2020). The multiple gRNA was cloned to target C1 and Rep genes

simultaneously. The resulted cottons plants did not display any CLCuD symptoms after 30 days of transformation and the results were further confirmed through PCR by extracting the RNA at different growth stages (Shafique et al., 2022). Shortly, prime editing represents a significant advancement in genome-editing technology, offering improved precision and efficiency compared to traditional methods (Lin et al., 2020). The development of various prime editing systems, from PE1 to PE3b, has enabled researchers to achieve precise modifications in both human cells and crop plants. As the technology continues to evolve, prime editing may become a powerful tool in the hands of cotton breeders for creating sustainable and resilient cotton varieties to meet the challenges of the future (Li et al., 2020).

6 Beyond genome editing

6.1 Next-generation sequencing technologies

Cotton has been extensively studied in the field of genomics to understand its genetic makeup and improve its agronomic traits. In the past, classical methods such as microarray analysis were used to study the genetics of plants, but they were costly, low-yielding, slow and labor-intensive (Hu et al., 2021). However, in the last two decades, high-throughput sequencing methods, collectively called Next-Generation Sequencing (NGS), have revolutionized plant genetics research, including cotton. NGS methods offer higher throughput, faster turnaround times, and reduced costs compared to traditional sequencing methods (Salk et al., 2018).

6.2 First generation technology

The first generation of sequencing technology, which emerged in the early 2000s, included chemical analysis and enzymatic sequencing methods. These methods relied on chain termination using 2',3'- dideoxynucleotides (ddNTPs) (Medžiūnė et al., 2022). First-generation sequencers, such as the ABI Prism, were developed to automate the sequencing process and could analyze 4 to 96 samples per single run (Kchouk et al., 2017). Although this technology produced high-quality sequencing, it had limitations in terms of laboriousness, slow methodology, and limited data output. It was primarily used for small genomes and large sequencing projects like the Human Genome Project (Collins et al., 2003). Mansoor et al. (1999) identified and characterized a new circular single-stranded (ss) DNA found in naturally infected cotton plants. Purified geminate particles from infected plants not only contain a component resembling DNA A but also a smaller, unrelated ssDNA termed DNA 1. DNA 1 was cloned from the double-stranded replicative form of viral DNA isolated from infected cotton plants. Blot hybridization using specific probes for CLCuV DNA or DNA 1 demonstrated the co-infection of naturally infected cotton plants in various geographical locations by both DNAs. DNA 1 was detected in viruliferous *Bemisia tabaci* and in laboratory-infected tobacco plants using *B. tabaci*, indicating its transmission by whiteflies. Sequence analysis revealed that DNA 1 is approximately half the size of CLCuV DNA, with no homology,

suggesting it is not a defective *Geminivirus* component. DNA 1 shares some homology with a genomic component of Nanoviridae, a family of DNA viruses typically transmitted by aphids or planthoppers. DNA 1 encodes a homologue of the nanovirus replication-associated protein (Rep) and can autonomously replicate in tobacco. The findings imply that a nanovirus-like DNA has acquired whitefly transmissibility through its association with a *Geminivirus*. This association may contribute to a mutually dependent relationship between two distinct DNA virus families with a similar replication strategy, potentially influencing the development of cotton leaf curl disease. Similarly, Sharma et al. (2005) used polymerase chain reaction (PCR) technique to detect viral DNA in samples stored for up to 3 days after collection, making it widely applicable for field diagnosis. In their study, the complete nucleotide sequence of the coat protein (CP) gene component of the Indian isolate of cotton leaf curl *Geminivirus* (CLCuV-HS2) was determined using CP-specific primers in PCR amplification. The infected cotton plants were obtained from fields in Haryana, India. Amino acid sequence comparison of the putative CP with other mono and bipartite *Geminiviruses* revealed a maximum identity of 97.3% with Pakistan cotton leaf curl virus (CLCuV-62). Additionally, a nuclear localization signal close to the N-terminal of the CP gene was identified. Moreover, Briddon et al. (2000) obtained full-length clones of Cotton Leaf Curl Virus (CLCuV), equivalent to the DNA A component of bipartite *Begomoviruses*. These CLCuV clones were found to be systemically infectious in both *Nicotiana benthamiana* and cotton plants. Interestingly, the infected plants did not display the typical symptoms associated with cotton leaf curl disease; instead, they exhibited mild leaf curling, yellowing, and some stunting. Despite efforts to identify a second genomic component, none was found. These results suggest that CLCuV alone may not be the sole cause of cotton leaf curl disease. However, the transmission of the disease by *Bemisia tabaci*, a whitefly, implies that the *Begomovirus* CLCuV might still play a role in disease transmission, even if it is not the primary causative agent.

6.3 Second generation technology

The advent of second-generation sequencing technology marked a significant milestone in the field of genomics. Unlike first-generation methods, second-generation or short-read sequencing techniques utilized PCR to amplify DNA fragments before sequencing. These techniques employed hybridization or synthesis-based methods to sequence millions of DNA fragments in parallel, enabling high-throughput sequencing (Adewale, 2020). Platforms like 454 Life Sciences Roche, SOLiD (Thermo-Fisher Scientific), Solexa (later acquired by Illumina), and Ion Torrent revolutionized the sequencing landscape. Illumina, the extensively employed technology for genotyping investigations in cotton, provides diverse choices for transcriptome and genome sequencing, delivering outputs spanning from gigabases to terabases (Meera Krishna et al., 2019). Patel (2016), detailed the *de novo* assembly process and characterization of transcriptome and genome data for an Asia I population. The research aimed to establish a genetic framework for insecticide discovery. Using the Roche 454 sequencing platform, the authors assembled a

comprehensive transcriptome of adult females from the Asia I population, generating 0.864 million reads and producing 29,418 contigs through CLC Genomics. Of these, 8,563 contigs were assigned putative functions. A draft genome of 828 Mbp for the Asia I population was constructed via *de novo* assembly of 206 million 250 bp paired-end reads from the Illumina MiSeq 2,500 platform using Platanus. This genome comprised 41,981 protein-coding genes (PCGs) and included 990 non-coding RNAs and repetitive elements (45.66%). The study examined 741 full-length genes, revealing larger intron sizes and repetitive elements contributing to the larger Asia I genome size, which explained challenges in genome assembly. Additionally, the research obtained one mitogenome and three endosymbiont genomes (Portiera, Wolbachia, and Arsenophonus) from the same sequence library, alongside the *B. tabaci* Asia I genome. The development of a genetic framework involved using *Drosophila* essential genes as a reference to identify orthologs in other insects, validated with ChEMBL targets. The presented 'omics' data serves as a comprehensive sequence resource for Asia I populations, illustrating the workflow for obtaining genetic information on hosts and their endosymbionts.

6.4 Third generation technology

Third-generation sequencing techniques, often referred to as single molecule sequencing (SMS), represent the latest advancement in sequencing technology. These methods do not require PCR amplification and can sequence single DNA molecules directly. The ability to sequence DNA molecules without amplification provides advantages such as reduced bias, improved accuracy, and the ability to detect modified bases. PacBio's Single Molecule Real-Time (SMRT) sequencing (Ardui et al., 2018), and Oxford Nanopore Technologies' nanopore sequencing (Weirather et al., 2017) are two prominent examples of third-generation sequencing platforms. These technologies have opened up new avenues for studying viral diversity, metagenomics, and epigenomics. Vij et al. (2022) presented a groundbreaking study on the introgression and mapping of Cotton Leaf Curl Disease (CLCuD) resistance from a 'synthetic cotton polyploid' to upland cotton. They developed a backcross population (synthetic polyploid/*Gossypium hirsutum* Acc. PIL 43/*G. hirsutum* Acc. PIL 43) to investigate the inheritance and mapping of CLCuD resistance. The study observed the dominance of CLCuD resistance over susceptibility, revealing the presence of two dominant genes conferring resistance to CLCuD. Molecular analysis using genotyping-by-sequencing identified chromosomes A01 and D07 as carrying one CLCuD resistance gene each. This research marks the first instance of successfully transferring and mapping CLCuD resistance from a synthetic polyploid to upland cotton. Furthermore, Hussain et al. (2023) conducted a sequencing study on the Mac7 accession, a source of *Gossypium hirsutum* known for its resistance against various biotic stresses. Through alignment with the *Gossypium hirsutum* 'TM-1' genome (AD1), they identified 4.7 million SNPs and 1.2 million InDels in the Mac7 genome. Gene ontology and metabolic pathway enrichment analyses indicated the involvement of these SNPs and InDels in processes such as nucleotide binding, secondary metabolite synthesis, and plant-pathogen interaction

pathways. Additionally, RNA-seq data from different tissues and qPCR expression profiling under Cotton Leaf Curl Disease (CLCuD) revealed the roles of individual genes in resistant and susceptible accessions. Notably, the expression of differential NLR genes was higher in resistant plants compared to susceptible ones. The resequencing results offer foundational data for identifying DNA resistance markers, potentially aiding marker-assisted breeding for the development of Mac7-derived resistance lines. Ahmed et al. (2021), utilized CIDER-seq (Circular DNA Enrichment Sequencing), a recently developed PCR-free virus enrichment sequencing method, in conjunction with Sanger sequencing to investigate the genetic diversity of the Cotton leaf curl disease (CLCuD) complex. The study identified a highly recombinant strain of Cotton leaf curl Multan virus and a recently evolved strain of Cotton leaf curl Multan betasatellite, both prevalent in major cotton-growing regions. Additionally, multiple species of alphasatellites were identified, including Mesta yellow vein mosaic alphasatellite (MeYVMA) observed for the first time in cotton. The study also employed real-time quantitative PCR to determine the relative abundance of the virus and associated satellites. This research represents the first study to characterize the CLCuD complex during its third epidemic.

7 Next-generation sequencing applications in cotton genomics

7.1 Genome sequencing and assembly

NGS technologies have displayed a significant role in sequencing and assembling the cotton genome. The availability of the cotton sequenced genome has facilitated the discoveries of genes responsible for important cotton traits, like disease resistance (CLCuD) and fiber quality. Both diploid and tetraploid cotton genomes have been sequenced using NGS technologies, providing valuable insights into the genetic basis of cotton traits. For example, Hussain et al. (2023), sequenced

G. hirsutum accessions named Mac7 (known to possess resistance against several biotic stresses) to identify the genetic bases associated with resistance against CLCuD. A total of 1.2 and 4.7 million InDels and SNPs were identified and their roles in resistance against CLCuD were accessed using metabolic pathway enrichment and gene ontology. Furthermore, RNA-seq data collected from different tissues at different growth stages identified individual genes associated with resistance and susceptibility in Mac7 against CLCuD (Hussain et al., 2023). Virus is famous for changing its strain rapidly and it became difficult to separate the newly developed strain from the previous one.

Genome sequencing technologies have now made it simple to separate the new strain one from the previous one and also detect the genes/alleles associated with the CLCuD. For example, Kumar et al. (2010) collected unknown leaf sample affected with CLCuD from two different sites of Rajasthan and DNA-A components of both samples were characterized and amplified. DNA-A of one sample consist of 2,759 nucleotides and the second sample consist of 2,751 nucleotides with both having total of six open reading frames (ORF's). Upon sequence characterization, the first sample

showed sequence similarity with CLCuV strain reported in Pakistan and the second sample showed sequence similarity with the local Rajasthan CLCuV strain. These sequence findings suggest a potential movement of certain strains of the CLCuV from Pakistan to India, or the concurrent existence of diverse isolates in similar geographical conditions. Whereas CLCuV-SG01 shared the highest nucleotide sequence similarity with CLCuV Rajasthan (Abohar), the V1 ORF (encoding coat protein) of SG01 displayed the highest nucleotide identity (100%) with CLCuV Multan (Bhatinda) and Abohar virus, albeit with variations in the AC1 region. The entire nucleotide sequence of SG01 exhibited an 86% similarity to CLCuV Multan virus. A similarity search unveiled significant disparities in AC1 and AV1 regions concerning DNA-A, indicative of a recombination-driven evolutionary history. Computational analysis through the Recombination Detection Program (RDP) supports this recombination hypothesis, suggesting the occurrence of recombination with other *Begomoviruses* within the AC1-ORF of CLCuV-SG01 and V1 ORF, as well as the AC1-ORF of CLCuV-SG02, alongside the intergenic noncoding region (Kumar et al., 2010).

7.2 Transcriptome analysis and gene expression profiling

RNA sequencing (RNA-seq) has become a popular approach for studying gene expression in cotton. By sequencing the transcriptome, researchers can identify differentially expressed genes under various conditions, such as stress, development, and tissue-specific expression (Berkowitz et al., 2021). RNA-seq has enabled the discovery of non-coding RNAs, alternative splicing events and novel transcripts in cotton (Qin et al., 2022; Singh et al., 2023). The commonly grown cotton species, *Gossypium hirsutum* is easily affected by CLCuD whereas; the diploid species *G. arboreum* naturally has some resistance (Hussain et al., 2023). However, the mechanism behind how CLCuD affects the genes expression profile of *G. arboreum* and how the disease interacts with it is unclear. Therefore, Naqvi et al. (2017) utilized RNA-Seq technique to look at how genes behave in *G. arboreum* when it is dealing with CLCuD. They infected *G. arboreum* plants with CLCuD by 600 grafting parts from infected *G. hirsutum* plants and studied the genes in both healthy and infected *G. arboreum* plants using RNA-seq with an Illumina HiSeq 2,500 machine. A total of 1,062 differentially expressed genes (DEGs) were identified of which 17 significant DEGs having possible role in disease resistance were studied for their expression profiles using qPCR. They also found genes that are related to fighting the disease and protecting the plant. Using the RNA-Seq data, they made a network of genes that work together. Within this network, they identified 50 hub genes, many of which help transport things inside the plant. These genes might be important for how *G. arboreum* defends itself against CLCuD. This study is a starting point for understanding how the virus and the plant interact and finding the key genes that help *G. arboreum* resist CLCuD (Naqvi et al., 2017). Furthermore, transcriptome analysis and gene expression profiling offer insights into how the plant reacts when CLCuD is transmitted through whiteflies in *G. hirsutum*. Naqvi et al. (2019) utilized RNA-Seq approach to delve into the variations in gene activity

within the susceptible *G. hirsutum* variety when it is infected by whiteflies carrying the CLCuD. To compare, they subjected both uninfected and CLCuD-infected plants to RNA-Seq analysis using the Illumina HiSeq 2,500 platform. They identified 468 differentially expressed genes (DEGs), consisting of 248 downregulated and 220 upregulated DEGs that play roles in responding to diseases and defending against pathogens. Ten of these genes were selected for further analysis using RT-qPCR on two susceptible cultivars, MNH 786 and Karishma. The consistent gene expression patterns across both cultivars affirmed the reliability of their transcriptome data, suggesting the applicability of their study's insights into the broader context of host susceptibility to CLCuD. They proceeded to conduct weighted gene co-expression network analysis, revealing six modules within the gene interaction networks. This analysis also pinpointed highly co-expressed genes, with 55 hub genes demonstrating co-expression with at least 50 other genes. Notably, many of these hub genes were found to be downregulated and involved in cellular processes. The diminished expression of these co-expressed genes indicated their potential role in favoring the virus and amplifying the plant's vulnerability to CLCuD. They also discuss the potential mechanisms underpinning the establishment of disease susceptibility. Altogether, their study presents a comprehensive analysis of gene activity changes in *G. hirsutum* when infected by whitefly transmitted CLCuD. This research contributes to a better understanding of the combined impact of whiteflies and the virus on their host, offering insights into significant *G. hirsutum* genes that play a role in its susceptibility to CLCuD (Naqvi et al., 2019). Transcriptome analysis and gene expression profiling have also been successfully applied to crops other than cotton, such as wheat (Lee et al., 2022; Xu et al., 2022), rice (Thomas et al., 2019; Kong et al., 2020), maize (Wang et al., 2019; Zhang et al., 2020), tomato (Wen et al., 2019; Riccini et al., 2021) and many others.

7.3 Identification of genetic variation

The creation of genetic diversity has always been the prime target of plant breeders to improve crop growth, development, and yield. Plant breeders use different techniques from hybridization to induce mutations (chemical or physical) to create genetic diversity in plants (Wang et al., 2020). In the case of CLCuD, inducing mutation via chemical or physical means to develop resistance against CLCuD in susceptible cotton genotypes is another modern breeding technique (Rehman et al., 2023). However, after inducing tolerance against CLCuD via mutation in susceptible line, detecting the mutated gene or single nucleotide polymorphisms (SNPs) associated with resistance is another objective of breeders which can only be detected after whole genome sequencing (Chen et al., 2022). NGS technologies have revolutionized the identification and characterization of genetic variation in cotton. By sequencing multiple cotton genotypes, researchers can detect SNPs, insertions, deletions, and structural variations. This information is crucial for marker-assisted selection (MAS) or marker-assisted breeding (MAB), genetic mapping, and understanding the complex traits genetic architecture in cotton (Kushanov et al., 2021). Molecular markers have shown important role in agriculture for cultivar and agronomical trait improvement such as detection of resistance genes

in plants. Numerous QTLs have been developed based on different experimental designs and genome mapping has been greatly improved and has provided a number of QTLs for major and minor crops including cotton (Zargar et al., 2015). Molecular markers linked with QTLs have been identified for various diseases in different crops. MAS has been successfully used in the development of resistant varieties. In cotton, QTLs related to disease are described as simple inherited traits. However, molecular breeding for disease resistance/tolerance is more specific, that is why a lot of work is needed to be done for locating QTLs linked with CLCuD resistance in cotton (Javed et al., 2019). In the recent years, genome map construction has gained an immense significance in MAS. MAS has been acquired as a molecular tool for the selection of desirable phenotypes using DNA markers such as RFLP, RAPD and AFLP on the basis of agriculturally important traits and the genes linked with tolerance against several abiotic and biotic factors (Jones et al., 1997; Khan et al., 2023). Use of MAS in plant breeding has opened new doors in crop improvement (Winter and Kahl, 1995).

Although, CLCuD has always been an important topic of discussion for botanists, yet exact opinion is still needed to be developed on the inheritance and spread of CLCuD (Mahmood et al., 2003). A study conducted by Rahman et al. was based on the collection of minor genes associated with CLCuV disease resistant material, selected by recurrent method (Rahman et al., 2005). Resistance is controlled by major genes, which are dominant and may be affected by the evolution of pathogens (Iqbal et al., 2003).

7.4 Epigenomics and DNA methylation analysis

Epigenetic modifications, such as DNA methylation, play a crucial role in regulating gene expression. NGS technologies have enabled the study of DNA methylation patterns in cotton, providing insights into epigenetic regulation of important agronomic traits. DNA methylation analysis can help understand the impact of environmental factors on gene expression and phenotype (He et al., 2022). DNA methylation in plants take place at cytosine residues in the context of CHH, CHG and CG (where H can be T, C or A) sequences (Feng et al., 2020). DNA methylation patterns can be heritable and can influence gene expression and phenotype. In cotton, DNA methylation has been shown to play a role in various developmental processes, stress responses, and disease resistance. To discover DNA methylation alterations during cotton domestication and evolution, researchers have developed single-base resolution methylomes from different cotton species and varieties. They found significant differences in DNA methylation patterns among diploid and tetraploid cottons, as well as between wild and cultivated cottons (Song et al., 2017). These DNA methylation changes were found to be associated with gene expression changes and may have contributed to the domestication and adaptation of cotton plants. Moreover, transposable elements (TEs) are usually linked with DNA methylation and can contribute to complexity of genome. In diploid cotton species, there are differences in TE content and DNA methylation patterns between different species (Song et al., 2017). The presence of TEs in genic regions can affect gene expression and evolution. The study of DNA methylation divergence among diploid

cotton species provides deep information in understanding the role of DNA methylation in genome evolution and speciation. DNA methylation can play a role in disease resistance in plants (Tirnaz and Batley, 2019). In cotton, DNA methylation changes have been associated with resistance to CLCuD. The betasatellite, a key component of the CLCuD complex, is involved in manipulating host cellular functions and suppressing gene silencing. DNA methylation changes in response to betasatellite infection may contribute to disease resistance by regulating gene expression and host defense responses. qPCR analysis has been used to evaluate the CLCuD-associated *Begomoviruses* titers and their impact on cotton yield. qPCR is a powerful method for quantifying the levels of virus and satellite DNA in infected plants. The results of qPCR analysis have unveiled a significant negative correlation between betasatellite titer and seed cotton yield (SCY) (Iqbal et al., 2023). This finding suggests that betasatellite titer is a critical factor in determining cotton yield and productivity.

7.5 Functional genomics

Functional genomics has emerged as a crucial scientific discipline in recent years, revolutionizing our understanding of plant biology and genetics improvement (Van Bel et al., 2022). In the case of cotton, functional genomics offers the potential to systematically exploit genetic resources, improve resistance against biotic stresses, enhance yield, and optimize germplasm for cultivation (Chen et al., 2007; Wei et al., 2022). However, several challenges persist in cotton functional genomics, including the long growth cycle, low transformation efficiency, large genome size and the lack of efficient molecular tools. Nonetheless, considerable advancements have been achieved in the development of genetically modified cotton varieties, particularly against resistance to CLCuD and other insects (Shafique et al., 2022). The sequencing of genomes plays a pivotal role in functional genomics. In 2007, the Cotton Genome Consortium strategized to sequence cotton genomes, prioritizing less-complicated diploid genomes that could be directly applied to tetraploid cotton (Wang et al., 2012). Consequently, the draft genome sequence of *G. raimondii*, a diploid species, was released in 2012, providing valuable insights into the larger 'A' diploid and 'AD' tetraploid cotton genomes (Li et al., 2014). Subsequently, the genome of *G. arboreum*, a supposed donor species for the A chromosome group in tetraploid cotton, was also sequenced (Watson et al., 2018). However, the precise species directing the development of tetraploid cotton currently remains unknown, highlighting the need for further genome sequencing to understand the biology and evolutionary history of CLCuD. Despite the availability of reference genome sequences for tetraploid and diploid cotton species, some discrepancies exist, potentially due to assembly errors. Therefore, efforts should be directed towards improving the quality control standards for genome assemblies and re-sequencing for accurate comparison and analysis of different cotton species. Vij et al. (2022) used one of the functional genomics approach (genotyping by sequencing) to detect the genes associated with the CLCuD in both upland cotton and backcross population (synthetic cotton polyploidy). They identified two potential genes on chromosomes D07 and

A01 that confer resistance to CLCuD (Vij et al., 2022). Similarly; Farooq et al. (2022), compared the CLCuD transmission behavior of different whitefly species (MEAM1 and Asia II 7) with different time frames using transcriptomics. Sequencing using Illumina technology unveiled, after 6 and 12 h of acquiring CLCuD, different sets of genes were differentially expressed in viruliferous (VF) whiteflies (MEAM1 and Asia II 7) compared to aviruliferous (AVF) whiteflies. The main categories of these DEGs were associated with transcription factors (TFs) of HTH-1 class in Asia II 7 and zf-C2H2 class in MEAM1. Interestingly, MEAM1 species showed more transcriptional changes and increased immune-related responses after CLCuD infection compared with Asia II 7. Both species had specific genes linked to antiviral innate immunity, lysosome function, autophagy/phagosome pathways, spliceosome, detoxification, signaling pathways, carbohydrate metabolism, transport, and protein digestion. Validation through RT-qPCR and RNA-seq showed consistent expression in 23 out of 28 selected genes. These findings provide insights into how *Begomoviruses* interact with whiteflies, affecting their transmission dynamics, and offer new targets for managing insect-transmitted plant viruses in a sustainable way (Farooq et al., 2022).

To facilitate the comprehensive study of cotton genomes, several functional genomics databases have been developed for the cotton research community. These databases provide access to crucial information such as genome sequences, map positions, protein expression and mRNA data, allelic variation and metabolic pathways. CottonGen (<https://www.cottongen.org>), is a curated web based intellectual database that offers convenient access to cotton genetic and genomic data. It includes germplasm resources, genes, trait loci, markers, genetic maps, unigenes from ESTs and annotated whole-genome sequences. Another significant database, Cotton Functional Genomic Database (CottonFGD; <https://cottonfgd.org>), provides access to transcriptome data, functional annotations, genome sequences, and genome re-sequencing data for sequenced *Gossypium* genomes. Additionally, databases like the Database for Co-expression Networks with Function Modules (ccNET; <http://structuralbiology.cau.edu.cn/gossypium/>), Join Genome Institute (JGI; <http://jgi.doe.gov>), Comparative Evolutionary Genomics of Cotton (<http://cottonrevolution.info/>), Platform of Functional Genomics Analysis in *Gossypium raimondii* (GraP; <http://structuralbiology.cau.edu.cn/GraP/about.html>), Cotton Genome Database (CottonDB; <http://www.cottondb.org>), and Cotton Genome Resource Database (CGRD; <http://cgrd.hzau.edu.cn/index.php>) offer valuable resources for in-depth analysis and exploration of cotton functional genomics.

7.6 Speed breeding

In the world of plant breeding, the traditional methods of crossbreeding and selection can be 757 time-consuming and labor-intensive. However, with the advent of speed breeding, plant breeders now have a powerful tool at their disposal to accelerate the generation of new plant varieties. Speed breeding involves manipulating environmental factors to promote rapid plant

growth and shorter generation times and development of tolerance against diseases like CLCuD (Jähne et al., 2020). Plant breeding has long been a slow and meticulous process, often taking years or even decades to develop new crop varieties with desirable traits. This time-consuming nature of traditional breeding techniques has posed challenges in addressing urgent agricultural needs, such as developing crops with high nutritional profile, higher yield, and resilience to biotic and abiotic factors. This is where speed breeding comes into play (Ghosh et al., 2018).

Speed breeding offers a way to shorten the breeding cycle by creating optimal growth conditions for plants. By manipulating factors such as light exposure, temperature, and humidity, breeders can accelerate plant growth and achieve multiple generations within a single year. This rapid generation turnover opens up new possibilities for crop improvement and allows breeders to respond more effectively to emerging challenges in agriculture (Chirugwi et al., 2019). To combat CLCuD and develop resistant cotton varieties, speed breeding has emerged as a promising technique. By accelerating the breeding cycle, breeders can quickly screen large populations of cotton plants for disease resistance and select the best-performing individuals for further breeding. This rapid selection process allows breeders to develop new cotton varieties with improved resistance to CLCuD in a shorter timeframe. So far, no research has reported the application of speed breeding against CLCuD in cotton. In addition to cotton, speed breeding has been used successfully in other cereal crops such as wheat and rice to improve disease resistance and final grain yield (Alahmad et al., 2018; Chimmili et al., 2022; Schoen et al., 2023). Given the successful use of speed breeding in other crops, cotton breeders and molecular biologists are now working on its use to improve biotic and abiotic stress tolerance in cotton.

One of the critical environmental factors in speed breeding is lighting. Plants have well-established light-sensitive receptor pathways that respond to specific wavelengths of light, triggering growth and development processes (Hickey et al., 2019). In speed breeding, breeders use specialized LED lamps to provide the necessary light spectrum for optimal plant growth and development. LED lamps offer several advantages over traditional lighting sources, such as fluorescent lamps. LED lamps can be precisely tuned to emit specific wavelengths of light; allowing breeders to customize the light spectrum based on the crop and desired traits (Gavhane et al., 2023). By manipulating the blue and red light wavelengths, breeders can induce early flowering in plants, effectively shortening the breeding cycle. Furthermore; Nexsel, a leading provider of speed breeding solutions, has been at the forefront of developing optimized lighting protocols for various crops, including cotton. With their specialized grow cabinets and a tailored lighting solution, Nexsel offers a comprehensive package for speed breeding in cotton. The Nexsel Speed Breeding Growth Chamber provides precise control over critical parameters such as temperature, humidity, CO₂ levels, lighting spectrum, light intensity, and photoperiod. This level of control allows breeders to create the ideal growth conditions for cotton plants, maximizing their potential for rapid growth and development. Nexsel's lighting solutions for speed breeding, such as the All-in-One Grow Light and the 3-in-1 Grow Light, offer customizable spectra and intensity options to meet the specific needs of cotton plants at different growth stages (Jois, 2021; Gavhane et al., 2023).

8 Current challenges and future directions

In recent years, advanced biotechnological and molecular breeding techniques have shown promise in the field of cotton research. Precision genome editing *via* multiplex genome editing and HDR have emerged as potential strategies for controlling CLCuD and improving cotton varieties. Additionally, the *de novo* domestication of wild species and the use of synthetic biology approaches offer new avenues for developing resistance against CLCuD and enhancing cotton productivity.

8.1 Precision genome editing via HDR in cotton

Precision genome editing through HDR holds great potential for accelerating cotton improvement and developing resistance against CLCuD. HDR allows for the precise replacement of existing alleles, offering a more targeted and efficient approach compared to traditional breeding methods. However, the application of HDR for particular allele or gene replacement in cotton remains challenging, and as of now, no stable lines with precise editing have been successfully obtained. To enhance HDR efficiency in cotton, several strategies used in other plant species and mammalian cells can be evaluated. One such strategy is the use of a modular RNA aptamer-streptavidin system to enrich the availability of donor repair templates (DRTs), which has shown significant improvements in HDR efficiency in human cells (Carlson-Severmer et al., 2017).

Another approach involves increasing the temporal and spatial colocalization of DRTs and the Cas9 protein, which can be achieved through the use of SNAP-tag technology. Covalently tethering single-stranded donor oligonucleotides (ssODNs) to the Cas9/guide RNA ribonucleoprotein complex has also been shown to increase HDR efficiency by bringing the DRTs closer to the DNA double-strand break (DSB) site (Savic et al., 2018). Furthermore, utilizing tissue-specific promoters, such as the egg cell- and early embryo-specific DD45 gene promoter, can increase the efficiency of HDR-mediated genome editing. This is particularly relevant in cotton, as HDR take place with more frequency in egg cells (Miki et al., 2018; Wolter et al., 2018). Additionally, the development of a tandem repeat-HDR (TR-HDR) strategy (Lu et al., 2020), which involves the random insertion of a modified DNA fragment, shows promise for seamless in-locus tagging and fragment replacement in other crops and could potentially be applied to cotton. While these techniques show potential, further research is needed to optimize and adapt them specifically for cotton. Continued efforts to improve HDR efficiency in cotton will pave the way for precise gene/allele replacement and the development of novel germplasm with resistance to CLCuD.

8.2 Multiplex genome editing in cotton

Many agriculturally important traits, including yield per plant, are governed by a complex 827 genetic network or multiple genes. In order to understand the functions of these genes more thoroughly,

expedite gene discovery, and facilitate breeding efforts, a proficient multiplex genome editing platform is essential. Multiplex genome editing enables the knockout of individual genes, combined genes, or specific gene regions, providing the means to dissect traits and support marker-assisted breeding. In cotton, the formulation of a proficient multiplex editing strategy is still in its early stages. However, insights from other plant species, such as wheat (Kumar et al., 2019) and *Arabidopsis* (Luo et al., 2021), can guide the optimization and systemic comparison of different vector components in cotton. For example, the use of robust constitutive promoters for the expression of Cas9 protein or guide RNA (gRNA), along with additional nuclear localization signals (NLSs), can contribute to the development of a more effective multiplex editing strategy.

Recent advancements in multiplex genome editing in wheat using a polycistronic tRNA strategy offer potential insights for cotton research. In an elite wheat variety, the simultaneous editing of multiple genes at up to 15 genomic loci was achieved by producing multiple mature single chimeric guide RNAs (sgRNAs) simultaneously. This approach utilized an expression cassette controlled by a rice Actin promoter and terminated by a Poly A sequence to enhance the transcripts stability (Luo et al., 2021). By adapting similar strategies to cotton, it may be possible to achieve efficient multiplex editing and accelerate gene discovery and breeding efforts. Expanding the horizons of multiplex editing in cotton can also involve the use of multiple Cas nucleases with different PAM requirements, like Cas12a, SaCas9, SpRY, SpG, xCas9 and SpCas9-NG (Zhang et al., 2019a; Jordan et al., 2021; Li et al., 2021). Additionally, multiplexing systems can be employed to target non-coding RNA regions and several other genetic regulatory elements, enabling the evaluation of their roles and further promoting cotton improvement against CLCuD. The development of a flexible and efficient CRISPR/Cas multiplexing system in cotton will highly facilitate fundamental biological research and translational breeding processes.

8.3 *De novo* domestication of wild species and site-directed artificial evolution of agriculturally important genes

Conventional breeding practices usually result in low stress resistance and loss of genetic diversity, posing challenges to crop stability and food security. To address these issues, the *de novo* domestication of wild species and site-directed artificial evolution of agriculturally important genes have emerged as potential approaches (Hickey et al., 2019). Through the comprehensive application of synthetic biology, genome editing, genomics and other cutting-edge technologies, wild species or landraces can be rapidly domesticated without disturbing the important traits (Eshed and Lippman, 2019). For instance, the introduction of traits of interest into stress-resilient wild tomato genotypes *via* multiplex genome editing resulted in offspring that displayed domesticated phenotypes while retaining salt tolerance and parental disease resistance (Li et al., 2018; Zsögön et al., 2018).

A rational approach for developing new crops varieties involves the *de novo* domestication of wild allotetraploid rice (*Oryza alta*, CCDD) has also been proposed. By editing several traits in *O. alta*

and successfully producing allotetraploid rice lines with targeted improvements in agronomically important genes, researchers have demonstrated the potential for *de novo* domestication in cotton relatives or landraces (Yu et al., 2021). Additionally, base editing techniques could permit the artificial evolution of agriculturally significant genes in existing crop varieties, leading to the development of new germplasm with high genetic diversity. While these approaches have not been documented in cotton to date, the lessons learned from other crop species can inform future research and applications. The exploration of wild cotton species and related plant genomes, along with the manipulation of diverse gene modules, will provide cotton breeders with novel genes and modules for crop improvement. Comparative genomic analysis and in-depth studies of *Gossypium* species will contribute to the development of novel germplasm and the improvement of cotton varieties against CLCuD.

8.4 Development of genotype-independent and in planta delivery strategies

CRISPR-based genome editing techniques rely on the delivery of foreign DNA into plant cells to enable the function and expression of the CRISPR/Cas reagents. In cotton, this has traditionally been achieved through *Agrobacterium*-mediated transformation or biolistic transformation (Ribeiro et al., 2003; Rajasekaran, 2012). However, these methods are usually limited by low plant regeneration efficiency and the limited number of transformable genotypes, particularly in modern cotton varieties. To overcome these limitations, the development of genotype-independent and *in planta* delivery strategies is essential. A genotype-independent approach has been implemented in maize by introducing the Bbm expression cassette into embryos under the control of the maize auxin-inducible promoter (Lowe et al., 2018). This approach allows for the direct germination of embryos into plants without a callus phase, enabling more uniform and efficient transformation. Synthetic biology approaches, which involve the design and construction of new biological parts and systems, also offer potential solutions for improving cotton transformation. By utilizing synthetic biology principles, it may be possible to redesign the cotton genome to enhance transformation efficiency and broaden the range of transformable genotypes.

9 Concluding remarks

The future of controlling and developing resistance in cotton against CLCuD lies in the 893 application of advanced biotechnological and molecular breeding techniques. Cotton employs a range of defense mechanisms against CLCuD, including resistance genes, RNA interference, and natural defense chemicals. While controlling whitefly populations through insecticides is the primary method, biotechnological and molecular breeding approaches offer the most promising solutions for inducing resistance. NGS technologies have transformed the field of cotton genomics, enabling researchers to unravel the complexities of the cotton genome and understand the

genetic basis of important traits. These technologies have opened up new avenues for studying gene expression, genetic variation, epigenomics, metagenomics, and functional genomics in cotton. With advancements in sequencing technologies and bioinformatics tools, the future of cotton genomics looks promising, paving the way for the development of improved cotton varieties with enhanced productivity, fiber quality, and disease resistance. Moreover, DNA methylation also plays a crucial role in cotton evolution, domestication, and disease resistance. Understanding the epigenetic regulation of CLCuD and the role of DNA methylation in cotton can provide valuable insights into the development of disease-resistant cotton varieties. The qPCR analysis of virus and satellite DNA levels can be a useful tool for assessing cotton yield and selecting resistant varieties for cultivation. Future research should focus on unraveling the complex interactions between DNA methylation, gene expression, disease resistance and genome editing in cotton to develop sustainable strategies for disease management in cotton cultivation. Precision genome editing *via* HDR holds promise for achieving specific allele/gene replacement and developing resistance against CLCuD. Similarly, multiplex genome editing offers opportunities for MAS, gene discovery and trait dissection. The newly domesticated wild species and site-directed artificial evolution of agriculturally important genes provide avenues for enhancing genetic diversity and developing novel germplasm. Additionally, the formulation of genotype-independent and *in planta* delivery strategies, along with the utilization of synthetic biology principles, can overcome the limitations of traditional transformation methods. Furthermore, speed breeding has also revolutionized the field of plant breeding, offering breeders a powerful tool to speed-up the development of new crop varieties. In the case of cotton, speed breeding holds immense promise for combating Cotton Leaf Curl Disease and addressing other agronomic challenges. By leveraging the advantages of specialized lighting solutions and controlled growth environments, breeders can expedite the breeding process, enhance disease resistance, improve fiber quality, increase yield, and develop climate-resilient cotton varieties. With companies like Nexsel leading the way in providing cutting-edge speed breeding solutions, the future of cotton breeding looks brighter than ever.

As these techniques continue to be optimized and adapted specifically for cotton, they have the potential to revolutionize cotton improvement and pave the way for sustainable cotton production. By developing cotton varieties with enhanced resistance against CLCuD, the devastating impacts of this disease can be mitigated, leading to increased yields and improved livelihoods for cotton farmers. The future of cotton research is bright, with biotechnological advancements driving innovation and unlocking the full potential of this important fiber crop.

Author contributions

SN: Conceptualization, Investigation, Writing—original draft. SR: Conceptualization, Writing—original draft, Writing—review and editing. TL: Writing—original draft. DT: Conceptualization, Formal Analysis, Supervision, Writing—review and editing. ZM: Investigation, Writing—original draft. KA: Formal Analysis,

Writing-review and editing. SM: Investigation, Writing-original draft. MT: Investigation, Writing-review and editing. NM: Investigation, Writing-review and editing. MK: Conceptualization, Data curation, Formal Analysis, Project administration, Resources, Supervision, Validation, Writing-review and editing. YL: Conceptualization, Formal Analysis, Investigation, Supervision, Validation, Visualization, Writing-review and 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 the Project of Sanya Yazhou Bay Science and Technology City (SCKJ-JYRC-2023-48).

References

- Adelewa, B. A. (2020). Will long-read sequencing technologies replace short-read sequencing technologies in the next 10 years? *Afr. J. laboratory Med.* 9 (1), 1–5. doi:10.4102/ajlm.v9i1.1340
- Aguiar, M. S., Quintela, E. D., Melo, C. L. P., Hungria, M., Faria, L. C., Wendland, A., et al. (2018). Agronomic performance and yield stability of the RNA interference-based bean golden mosaic virus-resistant common bean. *Crop Sci.* 58 (2), 579–591. doi:10.2135/cropsci2017.06.0355
- Ahmed, N., Amin, I., Zaidi, S. S. E. A., Rahman, S. U., Farooq, M., Fauquet, C. M., et al. (2021). Circular DNA enrichment sequencing reveals the viral/satellites genetic diversity associated with the third epidemic of cotton leaf curl disease. *Biol. Methods Protoc.* 6 (1), bpab005. doi:10.1093/biomethods/bpab005
- Ahmed, Z. (1999). Prospects and bottlenecks of cotton crop in Pakistan. *Pak Cotton Grow.* 3, 6–7.
- Alahmad, S., Dinglasan, E., Leung, K. M., Riaz, A., Derbal, N., Voss-Fels, K. P., et al. (2018). Speed breeding for multiple quantitative traits in durum wheat. *Plant methods* 14 (1), 36–15. doi:10.1186/s13007-018-0302-y
- Ali, A., Ahmed, Z., and Guo, Z. (2022). *Cotton growing in Pakistan, Bangladesh and Myanmar. Pest management in cotton: a global perspective*. Wallingford UK: CABI, 53–79.
- Ali, A., Zafar, M. M., Farooq, Z., Ahmed, S. R., Ijaz, A., Anwar, Z., et al. (2023). Breakthrough in CRISPR/Cas system: current and future directions and challenges. *Biotechnol. J.* 18, 2200642. doi:10.1002/biot.202200642
- Ali, H., Hussain, G., Hussain, S., Shahzad, A., Ahmad, S., Javeed, H., et al. (2015). Early sowing reduces cotton leaf curl virus occurrence and improves cotton productivity. *Cercetari Agronomice Moldova* 47 (4), 71–81. doi:10.1515/cerce-2015-0006
- Ali, I., Amin, I., Briddon, R. W., and Mansoor, S. (2013). Artificial microRNA-mediated resistance against the monopartite begomovirus Cotton leaf curl Burewala virus. *Virology J.* 10 (1), 231–238. doi:10.1186/1743-422X-10-231
- Amudha, J., Balasubramani, G., Malathi, V., Monga, D., Bansal, K., and Kranthi, K. (2010). Cotton transgenics with antisense AC1 gene for resistance against cotton leaf curl virus. *Electron. J. Plant Breed.* 1 (4), 360–369.
- Anzalone, A. V., Randolph, P. B., Davis, J. R., Sousa, A. A., Koblan, L. W., Levy, J. M., et al. (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* 576 (7785), 149–157. doi:10.1038/s41586-019-1711-4
- Ardui, S., Ameur, A., Vermeesch, J. R., and Hestand, M. S. (2018). Single molecule real-time (SMRT) sequencing comes of age: applications and utilities for medical diagnostics. *Nucleic acids Res.* 46 (5), 2159–2168. doi:10.1093/nar/gky066
- Arif, M., Farooq, S., Alasmari, A., Alshehri, M. A., Hashem, M., Alamri, S., et al. (2022). Molecular study of geminiviruses: complex biology, host-vector 970interactions, and increasing diversity. *J. King Saud University-Science* 34 (4), 102051. doi:10.1016/j.jksus.2022.102051
- Arora, S., and Chaudhary, B. (2021). Global expression dynamics and miRNA evolution profile govern floral/fiber architecture in the modern cotton (*Gossypium*). *Planta* 254 (3), 62. doi:10.1007/s00425-021-03711-3
- Ashraf, J., Zuo, D., Wang, Q., Malik, W., Zhang, Y., Abid, M. A., et al. (2018). Recent insights into cotton functional genomics: progress and future perspectives. *Plant Biotechnol. J.* 16 (3), 699–713. doi:10.1111/pbi.12856
- Atiq, M., Talib, M. Z., Rajput, N. A., Sahi, S. T., Khan, M. A., Usman, M., et al. (2023). NEW fangled tactics towards cotton leaf curl virus disease A review. *J. Nat. Fibers* 20 (2), 2217364. doi:10.1080/15440478.2023.2217364
- Azeem, H., Perveen, R., Tahir, M. N., Umar, U.-u.-d., Ölmez, F., and Ali, A. (2022). Prevalence, transmission and molecular characterization of Cotton leaf curl Multan virus infecting hollyhock plants in Pakistan. *Mol. Biol. Rep.* 49 (6), 5635–5644. doi:10.1007/s11033-022-07557-0
- Basu, U., Riaz Ahmed, S., Bhat, B. A., Anwar, Z., Ali, A., Ijaz, A., et al. (2023). A CRISPR way for accelerating cereal crop improvement Progress and challenges. *Front. Genet.* 13, 866976. doi:10.3389/fgene.2022.866976
- Berkowitz, O., Xu, Y., Liew, L. C., Wang, Y., Zhu, Y., Hurgobin, B., et al. (2021). RNA-seq analysis of laser microdissected *Arabidopsis thaliana* leaf epidermis, mesophyll and vasculature defines tissue-specific transcriptional responses to multiple stress treatments. *Plant J.* 107 (3), 938–955. doi:10.1111/tjp.15314
- Bharat, S. S., Li, S., Li, J., Yan, L., and Xia, L. (2020). Base editing in plants: current status and challenges. *Crop J.* 8 (3), 384–395. doi:10.1016/j.cj.2019.10.002
- Binyameen, B., Khan, Z., Khan, S. H., Ahmad, A., Munawar, N., Mubarik, M. S., et al. (2021). Using multiplexed CRISPR/Cas9 for suppression of cotton leaf curl virus. *Int. J. Mol. Sci.* 22 (22), 12543. doi:10.3390/ijms222212543
- Bourdenx, B., Bernard, A., Domergue, F., Pascal, S., Léger, A., Roby, D., et al. (2011). Overexpression of *Arabidopsis* ECERIFERUM1 promotes wax very-long-chain alkane biosynthesis and influences plant response to biotic and abiotic stresses. *Plant physiol.* 156 (1), 29–45. doi:10.1104/pp.111.172320
- Briddon, R., and Markham, P. (1994). Universal primers for the PCR amplification of dicot-infecting geminiviruses. *Mol. Biotechnol.* 1 (2), 202–205.
- Briddon, R. W., Mansoor, S., Bedford, I. D., Pinner, M. S., and Markham, P. G. (2000). Clones of cotton leaf curl geminivirus induce symptoms atypical of cotton leaf curl disease. *Virus Genes* 20 (1), 19–26. doi:10.1023/a:1008151921937
- Brown, J., and Czosnek, H. (2002). Whitefly transmission of plant viruses-III whitefly biology in relation to transmission. *Adv. Botanical Res.* 36, 71–74.
- Cao, Y., Zhou, H., Zhou, X., and Li, F. (2020). Control of plant viruses by CRISPR/Cas system-mediated adaptive immunity. *Front. Microbiol.* 11, 593700. doi:10.3389/fmicb.2020.593700
- Carlson-Stevermer, J., Abdeen, A. A., Kohlenberg, L., Goedland, M., Molugu, K., Lou, M., et al. (2017). Assembly of CRISPR ribonucleoproteins with biotinylated oligonucleotides via an RNA aptamer for precise gene editing. *Nat. Commun.* 8 (1), 1711. doi:10.1038/s41467-017-01875-9
- Chandran, D. (2015). Co-option of developmentally regulated plant SWEET transporters for pathogen nutrition and abiotic stress tolerance. *IUBMB life* 67 (7), 461–471. doi:10.1002/iub.1394
- Chauhan, P., Mehta, N., Chauhan, R., Kumar, A., Singh, H., Lal, M. K., et al. (2023). Utilization of primary and secondary biochemical compounds in cotton as diagnostic markers for measuring resistance to cotton leaf curl virus. *Front. Plant Sci.* 14, 1185337. doi:10.3389/fpls.2023.1185337
- Chen, C.-Y., Liu, Y.-Q., Song, W.-M., Chen, D.-Y., Chen, F.-Y., Chen, X.-Y., et al. (2019). An effector from cotton bollworm oral secretion impairs host plant defense signaling. *Proc. Natl. Acad. Sci.* 116 (28), 14331–14338. doi:10.1073/pnas.1905471116
- Chen, J.-Y., Huang, J.-Q., Li, N.-Y., Ma, X.-F., Wang, J.-L., Liu, C., et al. (2015). Genome-wide analysis of the gene families of resistance gene analogues in cotton and their response to Verticillium wilt. *BMC plant Biol.* 15, 1–15. doi:10.1186/s12870-015-0508-3
- Chen, Y., Gao, Y., Chen, P., Zhou, J., Zhang, C., Song, Z., et al. (2022). Genome-wide association study reveals novel quantitative trait loci and candidate genes of lint

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.

percentage in upland cotton based on the CottonSNP80K array. *Theor. Appl. Genet.* 135 (7), 2279–2295. doi:10.1007/s00122-022-04111-1

Chen, Z. J., Scheffler, B. E., Dennis, E., Triplett, B. A., Zhang, T., Guo, W., et al. (2007). Toward sequencing cotton (*Gossypium*) genomes. *Plant physiol.* 145 (4), 1303–1310. doi:10.1104/pp.107.107672

Chimmili, S. R., Kanneboina, S., Hanjagi, P. S., Ps, B., Sakhare, A. S., P, S., et al. (2022). "Integrating advanced molecular, genomic, and speed breeding methods for genetic improvement of stress tolerance in rice," in *Next-generation plant breeding approaches for stress resilience in cereal crops* (Singapore: Springer Nature Singapore), 263–283.

Chiurugwi, T., Kemp, S., Powell, W., and Hickey, L. T. (2019). Speed breeding orphan crops. *Theor. Appl. Genet.* 132, 607–616. doi:10.1007/s00122-018-3202-7

Collins, F. S., Morgan, M., and Patrinos, A. (2003). The Human Genome Project: lessons from large-scale biology. *Science* 300 (5617), 286–290. doi:10.1126/science.1084564

Devendran, R., Kumar, M., Ghosh, D., Yogindran, S., Karim, M. J., and Chakraborty, S. (2022). Capsicum-infecting begomoviruses as global pathogens: host–virus interplay, pathogenesis, and management. *Trends Microbiol.* 30 (2), 170–184. doi:10.1016/j.tim.2021.05.007

Diamos, A. G., Crawford, J. M., and Mason, H. S. (2019). Fine-tuning expression of begomoviral movement and nuclear shuttle proteins confers cell-to-cell movement to mastreviral replicons in *Nicotiana benthamiana* leaves. *J. General Virology* 100 (6), 1038–1051. doi:10.1099/jgv.0.001275

Dilip, M. (2016). Cotton leaf curl virus disease: an overview. *Agric. Res. J.* 53 (4), 466–474. doi:10.5958/2395-146x.2016.00093.4

Donnelly, R., and Gilligan, C. A. (2023). A new method for the analysis of access period experiments, illustrated with whitefly-borne cassava mosaic begomovirus. *PLoS Comput. Biol.* 19 (8), e1011291. doi:10.1371/journal.pcbi.1011291

Eshed, Y., and Lippman, Z. B. (2019). Revolutions in agriculture chart a course for targeted breeding of old and new crops. *Science* 366 (6466), eaax0025. doi:10.1126/science.aax0025

Faiz, H., and Abhinav, K. (2021). Development of immunity against cotton leaf curl virus using antisense ac4 gene. *Res. J. Biotechnol.* 16, 9.

Farooq, T., Lin, Q., She, X., Chen, T., Tang, Y., and He, Z. (2022). Comparative transcriptome profiling reveals a network of differentially expressed genes in Asia II 7 and MEAM1 whitefly cryptic species in response to early infection of Cotton leaf curl Multan virus. *Front. Microbiol.* 13, 1004513. doi:10.3389/fmicb.2022.1004513

Fauquet, C. M., and Stanley, J. (2003). Geminivirus classification and nomenclature: progress and problems. *Ann. Appl. Biol.* 142 (2), 165–189. doi:10.1111/j.1744-7348.2003.tb00241.x

Feng, S., Zhong, Z., Wang, M., and Jacobsen, S. E. (2020). Efficient and accurate determination of genome-wide DNA methylation patterns in *Arabidopsis thaliana* with enzymatic methyl sequencing. *Epigenetics chromatin* 13 (1), 42–17. doi:10.1186/s13072-020-00361-9

Fiallo-Olivé, E., and Navas-Castillo, J. (2023). Begomoviruses: what is the secret (s) of their success? *Trends Plant Sci.* 28, 715–727. doi:10.1016/j.tplants.2023.01.012

Figueiredo, A., Monteiro, F., and Sebastiana, M. (2014). Subtilisin-like proteases in plant–pathogen recognition and immune priming: a perspective. *Front. plant Sci.* 5, 739. doi:10.3389/fpls.2014.00739

Fondong, V. N. (2013). Geminivirus protein structure and function. *Mol. plant Pathol.* 14 (6), 635–649. doi:10.1111/mpp.12032

Fortes, I. M., Pérez-Padilla, V., Romero-Rodríguez, B., Fernández-Muñoz, R., Moyano, C., Castillo, A. G., et al. (2023). Begomovirus tomato leaf curl New Delhi virus is seedborne but not seed transmitted in melon. *Plant Dis.* 107 (2), 473–479. doi:10.1094/PDIS-09-21-1930-RE

Gaudelli, N. M., Komor, A. C., Rees, H. A., Packer, M. S., Badran, A. H., Bryson, D. I., et al. (2017). Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* 551 (7681), 464–471. doi:10.1038/nature24644

Gavhane, K. P., Hasan, M., Singh, D. K., Kumar, S. N., Sahoo, R. N., and Alam, W. (2023). Determination of optimal daily light integral (DLI) for indoor cultivation of iceberg lettuce in an indigenous vertical hydroponic system. *Sci. Rep.* 13 (1), 10923. doi:10.1038/s41598-023-36997-2

Ghosh, S., Watson, A., Gonzalez-Navarro, O. E., Ramirez-Gonzalez, R. H., Yanes, L., Mendoza-Suárez, M., et al. (2018). Speed breeding in growth chambers and glasshouses for crop breeding and model plant research. *Nat. Protoc.* 13 (12), 2944–2963. doi:10.1038/s41596-018-0072-z

Gnanasekaran, P., and Chakraborty, S. (2018). Biology of viral satellites and their role in pathogenesis. *Curr. Opin. virology* 33, 96–105. doi:10.1016/j.coviro.2018.08.002

Gnanasekaran, P., KishoreKumar, R., Bhattacharyya, D., Vinoth Kumar, R., and Chakraborty, S. (2019). Multifaceted role of geminivirus associated betasatellite in pathogenesis. *Mol. Plant Pathol.* 20 (7), 1019–1033. doi:10.1111/mpp.12800

Gould, J., and Northcote, D. (1986). Cell-cell recognition of host surfaces by pathogens the adsorption of maize (*Zea mays*) root mucilage by surfaces of pathogenic fungi. *Biochem. J.* 233 (2), 395–405. doi:10.1042/bj2330395

Gray, S. M., and Banerjee, N. (1999). Mechanisms of arthropod transmission of plant and animal viruses. *Microbiol. Mol. Biol. Rev.* 63 (1), 128–148. doi:10.1128/MMBR.63.1.128-148.1999

Guo, X., Wang, Y., Hou, Y., Zhou, Z., Sun, R., Qin, T., et al. (2022). Genome-wide dissection of the genetic basis for drought tolerance in *Gossypium hirsutum* L. races. *Front. Plant Sci.* 13, 876095. doi:10.3389/fpls.2022.876095

Hagan, A. K., Kloepper, J. W., Lawrence, K. S., Guertel, E. A., Smith, R. H., Shaw, J. N., et al. (2019). *Auburn university crops*. Cotton Research Report.

Harrison, B., Liu, Y., Khalid, S., Hameed, S., Otim-Nape, G., and Robinson, D. (1997). Detection and relationships of cotton leaf curl virus and allied whitefly-transmitted geminiviruses occurring in Pakistan. *Ann. Appl. Biol.* 130 (1), 61–75. doi:10.1111/j.1744-7348.1997.tb05783.x

Hasan, I., Rasul, S., Malik, T. H., Qureshi, M. K., Aslam, K., Shabir, G., et al. (2019). Present status of cotton leaf curl virus disease (CLCUVD): a major threat to cotton production. *Int. J. Cotton Res. Technol.* 1 (1), 1–13. doi:10.33865/ijcrt.001.01.0240

He, S., Zhang, Y., Wang, J., Wang, Y., Ji, F., Sun, L., et al. (2022). H3K4me2, H4K5ac and DNA methylation function in short-and long-term heat stress responses through affecting the expression of the stress-related genes in *G. hirsutum*. *Environ. Exp. Bot.* 194, 104699. doi:10.1016/j.envexpbot.2021.104699

Heigwer, F., Zhan, T., Breinig, M., Winter, J., Brügemann, D., Leible, S., et al. (2016). CRISPR library designer (CLD): software for multispecies design of single guide RNA libraries. *Genome Biol.* 17 (1), 55–10. doi:10.1186/s13059-016-0915-2

Hickey, L. T., Hafeez, A. N., Robinson, H., Jackson, S. A., Leal-Bertioli, S. C., Tester, M., et al. (2019). Breeding crops to feed 10 billion. *Nat. Biotechnol.* 37 (7), 744–754. doi:10.1038/s41587-019-0152-9

Hillary, V. E., and Caesar, S. A. (2022). Prime editing in plants and mammalian cells: mechanism, achievements, limitations, and future prospects. *BioEssays* 44 (9), 2200032. doi:10.1002/bies.202200032

Hu, T., Chitnis, N., Monos, D., and Dinh, A. (2021). Next-generation sequencing technologies: an overview. *Hum. Immunol.* 82 (11), 801–811. doi:10.1016/j.humimm.2021.02.012

Huang, G., Huang, J.-Q., Chen, X.-Y., and Zhu, Y.-X. (2021). Recent advances and future perspectives in cotton research. *Annu. Rev. plant Biol.* 72, 437–462. doi:10.1146/annurev-arplant-080720-113241

Hussain, A., Farooq, M., Naqvi, R. Z., Aslam, M. Q., Siddiqui, H. A., Amin, I., et al. (2023). Whole-genome resequencing deciphers new insight into genetic diversity and signatures of resistance in cultivated cotton *Gossypium hirsutum*. *Mol. Biotechnol.* 65 (1), 34–51. doi:10.1007/s12033-022-00527-8

Hussain, T., and Ali, M. (1975). *Review of cotton diseases of Pakistan*. Pakistan cottons.

Ibrahim, Y., Abdelhamed, M. F., and Ibrahim, F. M. (2023). Impact of Foliar Application of Some Growth stimulants on the vegetative growth and the Essential oil characters of *Ocimum basilicum* var. *thyrsiflorum* (L.). *Egypt. J. Chem.* 66 (4), 435–447. doi:10.21608/EJCHEM.2022.173276.7163

Iqbal, M., Chang, M. A., and Mahmood, A. (2003). *Inheritance of response to cotton leaf curl virus (CLCuV) infection in cotton*. Pakistan: Asian Journal of Plant Sciences. (Pakistan).

Iqbal, Z., Sattar, M. N., Kvarnheden, A., Mansoor, S., and Briddon, R. W. (2012). Effects of the mutation of selected genes of Cotton leaf curl Kokhran virus on infectivity, symptoms and the maintenance of Cotton leaf curl Multan betasatellite. *Virus Res.* 169 (1), 107–116. doi:10.1016/j.virusres.2012.07.016

Iqbal, Z., Shafiq, M., Ali, S., Mahmood, M. A., Siddiqui, H. A., Amin, I., et al. (2023). qPCR assay as a tool for examining cotton resistance to the virus complex causing CLCuD: yield loss inversely correlates with betasatellite, not virus, DNA titer. *Plants* 12 (14), 2645. doi:10.3390/plants12142645

Ismayil, A., Yang, M., Haxim, Y., Wang, Y., Li, J., Han, L., et al. (2020). Cotton leaf curl multan virus βC1 protein induces autophagy by disrupting the interaction of autophagy-related protein 3 with glyceraldehyde-3-phosphate dehydrogenases. *Plant Cell* 32 (4), 1124–1135. doi:10.1105/tpc.19.00759

Jähne, F., Hahn, V., Würschum, T., and Leiser, W. L. (2020). Speed breeding short-day crops by LED-controlled light schemes. *Theor. Appl. Genet.* 133 (8), 2335–2342. doi:10.1007/s00122-020-03601-4

Jain, H., Chahal, S., Singh, I., Sain, S. K., and Siwach, P. (2023). The rising threat of geminiviruses: molecular insights into the disease mechanism and mitigation strategies. *Mol. Biol. Rep.* 50 (4), 3835–3848. doi:10.1007/s11033-023-08266-y

Jastrzębska, M., Wachowska, U., and Kostrzewska, M. K. (2020). Pathogenic and non-pathogenic fungal communities in wheat grain as influenced by recycled phosphorus fertilizers: a case study. *Agriculture* 10 (6), 239. doi:10.3390/agriculture10060239

Javed, M., Hussain, S., and Baber, M. (2019). Role of QTL mapping to circumscribe various diseases in different crops with special emphasis on cotton. *J. Genet. Mol. Biol.* 3 (1). doi:10.35841/genetics-molecular-biology.3.23-33

Ji, X., Zhang, H., Zhang, Y., Wang, Y., and Gao, C. (2015). Establishing a CRISPR–Cas-like immune system conferring DNA virus resistance in plants. *Nat. Plants* 1 (10), 15144–4. doi:10.1038/nplants.2015.144

- Jois, S. (2021). Morphological characteristics of hydroponically grown lettuce (*Lactuca sativa*) treated with Pranik Agriculture. *Egypt. J. Agric. Res.* 99 (4), 391–396. doi:10.21608/EJAR.2021.98969.1159
- Jones, N., Ougham, H., and Thomas, H. (1997). Markers and mapping: we are all geneticists now. *New Phytologist* 137 (1), 165–177. doi:10.1046/j.1469-8137.1997.00826.x
- Jordan, W. T., Currie, S., and Schmitz, R. J. (2021). Multiplex genome editing in *Arabidopsis thaliana* using Mb3Cas12a. *Plant Direct* 5 (9), e344. doi:10.1002/pld3.344
- Joußen, N., Agnolet, S., Lorenz, S., Schöne, S. E., Ellinger, R., Schneider, B., et al. (2012). Resistance of Australian *Helicoverpa armigera* to fenvalerate is due to the chimeric P450 enzyme CYP337B3. *Proc. Natl. Acad. Sci.* 109 (38), 15206–15211. doi:10.1073/pnas.1202047109
- Karlson, C. K. S., Mohd-Noor, S. N., Nolte, N., and Tan, B. C. (2021). CRISPR/dCas9-based systems: mechanisms and applications in plant sciences. *Plants* 10 (10), 2055. doi:10.3390/plants10102055
- Karthik, K., Hada, A., Bajpai, A., Patil, B. L., Paraselli, B., Rao, U., et al. (2023). A novel tasi RNA-based micro RNA-induced gene silencing strategy to tackle multiple pests and pathogens in cotton (*Gossypium hirsutum* L.). *Planta* 257 (1), 20. doi:10.1007/s00425-022-04055-2
- Katiyar, A., Smita, S., Lenka, S. K., Rajwanshi, R., Chinnusamy, V., and Bansal, K. C. (2012). Genome-wide classification and expression analysis of MYB transcription factor families in rice and *Arabidopsis*. *BMC genomics* 13, 1–19. doi:10.1186/1471-2164-13-544
- Kchouk, M., Gibrat, J.-F., and Elloumi, M. (2017). Generations of sequencing technologies: from first to next generation. *Biol. Med.* 9 (3). doi:10.4172/0974-8369.1000395
- Khan, M. A. U., Shahid, A. A., Rao, A. Q., Shahid, N., Latif, A., ud Din, S., et al. (2015). Defense strategies of cotton against whitefly transmitted CLCuV and Begomoviruses. *Adv. Life Sci.* 2 (2), 58–66.
- Khan, M. K. R., Ditta, A., Wang, B., Fang, L., Anwar, Z., Ijaz, A., et al. (2023). “The intervention of multi-omics approaches for developing abiotic stress resistance in cotton crop under climate change,” in *Sustainable agriculture in the era of the OMICS revolution* (Springer), 37–82.
- Khan, S., Mahmood, M., Rahman, S., Rizvi, F., and Ahmad, A. (2020). Evaluation of the CRISPR/Cas9 system for the development of resistance against Cotton leaf curl virus in model plants. *Plant Prot. Sci.* 56 (3), 154–162. doi:10.17221/105/2019-pps
- Khan, Z. A., Kumar, R., and Dasgupta, I. (2022). CRISPR/Cas-mediated resistance against viruses in plants. *Int. J. Mol. Sci.* 23 (4), 2303. doi:10.3390/ijms23042303
- Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A., and Liu, D. R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* 533 (7603), 420–424. doi:10.1038/nature17946
- Kong, W., Zhang, C., Qiang, Y., Zhong, H., Zhao, G., and Li, Y. (2020). Integrated RNA-seq analysis and Meta-QTLs mapping provide insights into cold stress response in rice seedling roots. *Int. J. Mol. Sci.* 21 (13), 4615. doi:10.3390/ijms21134615
- Kumar, R., Kaur, A., Pandey, A., Mamrutha, H., and Singh, G. (2019). CRISPR-based genome editing in wheat: a comprehensive review and future prospects. *Mol. Biol. Rep.* 46 (3), 3557–3569.
- Kumar, S., Kanakachari, M., Gurusamy, D., Kumar, K., Narayanasamy, P., Kethireddy Venkata, P., et al. (2016). Genome-wide transcriptomic and proteomic analyses of bollworm-infested developing cotton bolls revealed the genes and pathways involved in the insect pest defence mechanism. *Plant Biotechnol. J.* 14 (6), 1438–1455. doi:10.1111/pbi.12508
- Kundu, P., Bera, P., Mishra, S., and Vadassery, J. (2023). *Regulatory role of phytohormones in the interaction of plants with insect herbivores*. Plant Hormones in Crop Improvement, Elsevier, 41–64.
- Kushanov, F. N., Turaev, O. S., Ernazarova, D. K., Gapparov, B. M., Oripova, B. B., Kudratova, M. K., et al. (2021). Genetic diversity, QTL mapping, and marker-assisted selection technology in cotton (*Gossypium* spp.). *Front. plant Sci.* 12, 779386. doi:10.3389/fpls.2021.779386
- Lee, M. H., Kim, K.-M., Sang, W.-G., Kang, C.-S., and Choi, C. (2022). Comparison of gene expression changes in three wheat varieties with different susceptibilities to heat stress using RNA-seq analysis. *Int. J. Mol. Sci.* 23 (18), 10734. doi:10.3390/ijms231810734
- Leke, W. N., Mignouna, D. B., Brown, J. K., and Kvarnheden, A. (2015). Begomovirus disease complex: emerging threat to vegetable production systems of West and Central Africa. *Agric. Food Secur.* 4 (1), 1–14. doi:10.1186/s40066-014-0020-2
- Li, F., Fan, G., Wang, K., Sun, F., Yuan, Y., Song, G., et al. (2014). Genome sequence of the cultivated cotton *Gossypium arboreum*. *Nat. Genet.* 46 (6), 567–572. doi:10.1038/ng.2987
- Li, J., Li, H., Chen, J., Yan, L., and Xia, L. (2020). Toward precision genome editing in crop plants. *Mol. Plant* 13 (6), 811–813. doi:10.1016/j.molp.2020.04.008
- Li, J., Zhu, L., Hull, J. J., Liang, S., Daniell, H., Jin, S., et al. (2016). Transcriptome analysis reveals a comprehensive insect resistance response mechanism in cotton to infestation by the phloem feeding insect *Bemisia tabaci* (whitefly). *Plant Biotechnol. J.* 14 (10), 1956–1975. doi:10.1111/pbi.12554
- Li, S., Zhang, C., Li, J., Yan, L., Wang, N., and Xia, L. (2021). Present and future prospects for wheat improvement through genome editing and advanced technologies. *Plant Commun.* 2 (4), 100211. doi:10.1016/j.xplc.2021.100211
- Li, T., Yang, X., Yu, Y., Si, X., Zhai, X., Zhang, H., et al. (2018). Domestication of wild tomato is accelerated by genome editing. *Nat. Biotechnol.* 36 (12), 1160–1163. doi:10.1038/nbt.4273
- Li, X., Schuler, M. A., and Berenbaum, M. R. (2002). Jasmonate and salicylate induce expression of herbivore cytochrome P450 genes. *Nature* 419 (6908), 712–715. doi:10.1038/nature01003
- Lim, H.-S., Nam, J., Seo, E.-Y., Nam, M., Vaira, A. M., Bae, H., et al. (2014). The coat protein of Alternanthera mosaic virus is the elicitor of a temperature-sensitive systemic necrosis in *Nicotiana benthamiana*, and interacts with a host boron transporter protein. *Virology* 452, 264–278. doi:10.1016/j.virol.2014.01.021
- Lin, Q., Zong, Y., Xue, C., Wang, S., Jin, S., Zhu, Z., et al. (2020). Prime genome editing in rice and wheat. *Nat. Biotechnol.* 38 (5), 582–585. doi:10.1038/s41587-020-0455-x
- Liu, S.-C., Feng, Y.-L., Sun, X.-N., Chen, R.-D., Liu, Q., Xiao, J.-J., et al. (2022). Target residence of Cas9-sgRNA influences DNA double-strand break repair pathway choices in CRISPR/Cas9 genome editing. *Genome Biol.* 23 (1), 165. doi:10.1186/s13059-022-02736-5
- Liu, Y., Robinson, D. J., and Harrison, B. D. (1998). Defective forms of cotton leaf curl virus DNA-A that have different combinations of sequence deletion, duplication, inversion and rearrangement. *J. General Virology* 79 (6), 1501–1508. doi:10.1099/0022-1317-79-6-1501
- Lowe, K., La Rota, M., Hoerster, G., Hastings, C., Wang, N., Chamberlin, M., et al. (2018). Rapid genotype “independent” *Zea mays* L.(maize) transformation via direct somatic embryogenesis. *Vitro Cell. Dev. Biology-Plant* 54, 240–252. doi:10.1007/s11627-018-9905-2
- Lu, G., Wang, Z., Xu, F., Pan, Y. B., Grisham, M. P., and Xu, L. (2021). Sugarcane mosaic disease: characteristics, identification and control. *Microorganisms* 9 (9), 1984. doi:10.3390/microorganisms9091984
- Lu, H., Cui, X., Liu, Z., Liu, Y., Wang, X., Zhou, Z., et al. (2018). Discovery and annotation of a novel transposable element family in *Gossypium*. *BMC plant Biol.* 18, 307–310. doi:10.1186/s12870-018-1519-7
- Lu, Y., Tian, Y., Shen, R., Yao, Q., Wang, M., Chen, M., et al. (2020). Targeted, efficient sequence insertion and replacement in rice. *Nat. Biotechnol.* 38 (12), 1402–1407. doi:10.1038/s41587-020-0581-5
- Luo, J., Li, S., Xu, J., Yan, L., Ma, Y., and Xia, L. (2021). Pyramiding favorable alleles in an elite wheat variety in one generation by CRISPR-Cas9-mediated multiplex gene editing. *Mol. Plant* 14 (6), 847–850. doi:10.1016/j.molp.2021.03.024
- Mahmood, T., Arshad, M., Gill, M. I., Mahmood, H. T., Tahir, M., and Hussain, S. (2003). Burewala strain of cotton leaf curl virus: a threat to CLCuV cotton resistant varieties. *Asian J. Plant Sci.* 2, 968–970. doi:10.3923/ajps.2003.968.970
- Majid, M. U., Sher, Z., Rashid, B., Ali, Q., Sarwar, M. B., Hassan, S., et al. (2020). Role of leaf epicuticular wax load and composition against whitefly population and cotton leaf curl virus in different cotton varieties. *Cytol. Genet.* 54, 472–486. doi:10.3103/s009545272005014x
- Mansoor, S., Bedford, I., Pinner, M., Stanley, J., and Markham, P. (1993). A whitefly-transmitted geminivirus associated with cotton leaf curl disease in Pakistan. *Pak. Botanical Soc. DEPT Bot. UNIV KARACHI, 32 KARACHI, Pak.* 25, 105–107.
- Mansoor, S., Bridson, R. W., Zafar, Y., and Stanley, J. (2003). Geminivirus disease complexes: an emerging threat. *Trends plant Sci.* 8 (3), 128–134. doi:10.1016/S1360-1385(03)00007-4
- Mansoor, S., Khan, S. H., Bashir, A., Saeed, M., Zafar, Y., Malik, K. A., et al. (1999). Identification of a novel circular single-stranded DNA associated with cotton leaf curl disease in Pakistan. *Virology* 259 (1), 190–199. doi:10.1006/viro.1999.9766
- Mao, Y.-B., Cai, W.-J., Wang, J.-W., Hong, G.-J., Tao, X.-Y., Wang, L.-J., et al. (2007). Silencing a cotton bollworm P450 monooxygenase gene by plant-mediated RNAi impairs larval tolerance of gossypol. *Nat. Biotechnol.* 25 (11), 1307–1313. doi:10.1038/nbt1352
- Marino, G., and Funk, C. (2012). Matrix metalloproteinases in plants: a brief overview. *Physiol. Plant.* 145 (1), 196–202. doi:10.1111/j.1399-3054.2011.01544.x
- Marwal, A., Sahu, A. K., and Gaur, R. K. (2014). “Transmission and host interaction of Geminivirus in weeds,” in *Plant virus–host interaction* (Academic Press), 143–161.
- Materatski, P., Jones, S., Patanita, M., Campos, M. D., Dias, A. B., Félix, M. D. R., et al. (2021). A bipartite geminivirus with a highly divergent genomic organization identified in olive trees may represent a novel evolutionary direction in the family Geminiviridae. *Viruses* 13 (10), 2035. doi:10.3390/v13102035
- Matsumoto, T., Lian, H.-L., Su, W.-A., Tanaka, D., Liu, C. w., Iwasaki, I., et al. (2009). Role of the aquaporin PIP1 subfamily in the chilling tolerance of rice. *Plant Cell physiology* 50 (2), 216–229. doi:10.1093/pcp/pcn190
- Medžiūnė, J., Kapustina, Ž., Žeimytė, S., Jakubovska, J., Sindikevičienė, R., Čikotienė, I., et al. (2022). Advanced preparation of fragment libraries enabled by oligonucleotide-modified 2', 3'-dideoxynucleotides. *Commun. Chem.* 5 (1), 34. doi:10.1038/s42004-022-00649-9
- Meera Krishna, B., Khan, M. A., and Khan, S. T. (2019). Next-generation sequencing(NGS) platforms: an exciting era of genome sequence analysis. *Microb. Genomics Sustain. Agroecosyst.* 2, 89–109. doi:10.1007/978-981-32-9860-6_6

- Meng, D., Walsh, M., and Fricke, W. (2016). Rapid changes in root hydraulic conductivity and aquaporin expression in rice (*Oryza sativa* L.) in response to shoot removal—xylem tension as a possible signal. *Ann. Bot.* 118 (4), 809–819. doi:10.1093/aob/mcw150
- 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 (1), 1967. doi:10.1038/s41467-018-04416-0
- Monga, D., and Sain, S. K. (2021). Incidence and severity of cotton leaf curl virus disease on different BG II hybrids and its effect on the yield and quality of cotton crop. *J. Environ. Biol.* 42 (1), 90–98. doi:10.22438/jeb/42/1/mrn-1296
- Montesinos, L., Bundó, M., Badosa, E., San Segundo, B., Coca, M., and Montesinos, E. (2017). Production of BP178, a derivative of the synthetic antibacterial peptide BP100, in the rice seed endosperm. *BMC plant Biol.* 17 (1), 63–14. doi:10.1186/s12870-017-1011-9
- Mubarik, M. S., Wang, X., Khan, S. H., Ahmad, A., Khan, Z., Amjid, M. W., et al. (2021). Engineering broad-spectrum resistance to cotton leaf curl disease by CRISPR-Cas9 based multiplex editing in plants. *GM Crops Food* 12 (2), 647–658. doi:10.1080/21645698.2021.1938488
- Musser, R. O., Hum-Musser, S. M., Eichenseer, H., Peiffer, M., Ervin, G., Murphy, J. B., et al. (2002). Herbivory: caterpillar saliva beats plant defences. *Nature* 416 (6881), 599–600. doi:10.1038/416599a
- Naqvi, R. Z., Zaidi, S. S.-e.-A., Akhtar, K. P., Strickler, S., Woldemariam, M., Mishra, B., et al. (2017). Transcriptomics reveals multiple resistance mechanisms against cotton leaf curl disease in a naturally immune cotton species, *Gossypium arboreum*. *Sci. Rep.* 7 (1), 15880. doi:10.1038/s41598-017-15963-9
- Naqvi, R. Z., Zaidi, S. S.-e.-A., Mukhtar, M. S., Amin, I., Mishra, B., Strickler, S., et al. (2019). Transcriptomic analysis of cultivated cotton *Gossypium hirsutum* provides insights into host responses upon whitefly-mediated transmission of cotton leaf curl disease. *PLoS one* 14 (2), e0210011. doi:10.1371/journal.pone.0210011
- Nogai, V. K., Singh, V., and Meghwal, R. R. (2014). Effect of Cotton leaf curl virus infected plants on the biology of the whitefly, *Bemisia tabaci* (Hemiptera: aleyrodidae): vector–virus mutualism. *Phytoparasitica* 42 (5), 619–625. doi:10.1007/s12600-014-0402-9
- Noueiry, A. O., Lucas, W. J., and Gilbertson, R. L. (1994). Two proteins of a plant DNA virus coordinate nuclear and plasmodesmal transport. *Cell* 76 (5), 925–932. doi:10.1016/0092-8674(94)90366-2
- Nurimanguli, A. I. N. I., Wu, Y. L., Pan, Z. Y., An, Q. S., Shui, G. L., Shao, P. X., et al. (2023). Cotton ethylene response factor GhERF91 involved in the defense against *Verticillium dahliae*. *J. Integr. Agric.* doi:10.1016/j.jia.2023.07.022
- Ouattara, A., Tiendrébéogo, F., Becker, N., Urbino, C., Thébaud, G., Hoareau, M., et al. (2022). Synergy between an emerging monopartite begomovirus and a DNA-B component. *Sci. Rep.* 12 (1), 695. doi:10.1038/s41598-021-03957-7
- Park, C.-J., and Seo, Y.-S. (2015). Heat shock proteins: a review of the molecular chaperones for plant immunity. *plant pathology J.* 31 (4), 323–333. doi:10.5423/PPJ.RW.08.2015.0150
- Patel, M. V. (2016). *De novo* assembly of transcriptome and genome data from the Asia I mtCOI genetic clade of *Bemisia tabaci*. Doctoral dissertation. University of Greenwich.
- Prasad, P., Khatoon, U., Verma, R. K., Sawant, S. V., and Bag, S. K. (2022). Data mining of transcriptional biomarkers at different cotton fiber developmental stages. *Funct. Integr. Genomics* 22 (5), 989–1002. doi:10.1007/s10142-022-00878-0
- Qadir, R., Khan, Z. A., Monga, D., and Khan, J. A. (2019). Diversity and recombination analysis of Cotton leaf curl Multan virus: a highly emerging begomovirus in northern India. *BMC genomics* 20 (1), 1–13. doi:10.1186/s12864-019-5640-2
- Qazi, J., Amin, I., Mansoor, S., Iqbal, M. J., and Briddon, R. W. (2007). Contribution of the satellite encoded gene betaC1 to cotton leaf curl disease symptoms. *Virus Res.* 128 (1–), 135–139. doi:10.1016/j.virusres.2007.04.002
- Qin, L., Li, J., Wang, Q., Xu, Z., Sun, L., Alariqi, M., et al. (2020). High-efficient and precise base editing of C● G to T● A in the allotetraploid cotton (*Gossypium hirsutum*) genome using a modified CRISPR/Cas9 system. *Plant Biotechnol. J.* 18 (1), 45–56. doi:10.1111/pbi.13168
- Qin, Y., Sun, M., Li, W., Xu, M., Shao, L., Liu, Y., et al. (2022). Single-cell RNA-seq reveals fate determination control of an individual fibre cell initiation in cotton (*Gossypium hirsutum*). *Plant Biotechnol. J.* 20 (12), 2372–2388. doi:10.1111/pbi.13918
- Rahman, M., Hussain, D., Malik, T., and Zafar, Y. (2005). Genetics of resistance to cotton leaf curl disease in *Gossypium hirsutum*. *Plant pathol.* 54 (6), 764–772. doi:10.1111/j.1365-3059.2005.01280.x
- Rajasekaran, K. (2012). *Biolistic transformation of cotton embryogenic cell suspension cultures*. Transgenic Cotton: Methods and Protocols, Springer, 59–70.
- Rasool, G., Yousaf, S., Ammara, U. e., Iqbal, A., Saeed, M., Amin, I., et al. (2021). Transgenic expression of synthetic coat protein and synthetic replication associated protein produces mild symptoms and reduce begomovirus-betasatellite accumulation in *Nicotiana benthamiana*. *Front. Agron.* 3, 676820. doi:10.3389/fagro.2021.676820
- Rehman, A., Almas, H. I., Qayyum, A., Li, H., Peng, Z., Qin, G., et al. (2023). *Mutation breeding in cotton. Biotechnologies and genetics in plant mutation breeding*. Apple Academic Press, 23–51.
- Ribeiro, S., Ambrozewicz, L., Avila, A., Bezerra, I., Calegario, R., Fernandes, J., et al. (2003). Distribution and genetic diversity of tomato-infecting begomoviruses in Brazil. *Archives virology* 148 (2), 281–295. doi:10.1007/s00705-002-0917-0
- Riccini, A., Picarella, M., De Angelis, F., and Mazzucato, A. (2021). Bulk RNA-Seq analysis to dissect the regulation of stigma position in tomato. *Plant Mol. Biol.* 105, 263–285. doi:10.1007/s11103-020-01086-9
- Rosario, K., Duffy, S., and Breitbart, M. (2012). A field guide to eukaryotic circular single-stranded DNA viruses: insights gained from metagenomics. *Archives virology* 157 (10), 1851–1871. doi:10.1007/s00705-012-1391-y
- Roumagnac, P., Lett, J.-M., Fiallo-Olivé, E., Navas-Castillo, J., Zerbini, F. M., Martin, D. P., et al. (2021). Establishment of five new genera in the family Geminiviridae: citlodavirus, maldovirus, mulcrilevirus, opunvirus, and topilevirus. *Archives virology* 167, 695–710. doi:10.1007/s00705-021-05309-2
- Saakre, M., Jaiswal, S., Rathinam, M., Raman, K. V., Tilgam, J., Paul, K., et al. (2023). Host-delivered RNA interference for durable pest resistance in plants: advanced methods, challenges, and applications. *Mol. Biotechnol.*, 1–20. doi:10.1007/s12033-023-00833-9
- Sahu, P. P., Sharma, N., Puranik, S., Muthamilarasan, M., and Prasad, M. (2014). Involvement of host regulatory pathways during geminivirus infection: a novel platform for generating durable resistance. *Funct. Integr. genomics* 14 (1), 47–58. doi:10.1007/s10142-013-0346-z
- Salk, J. J., Schmitt, M. W., and Loeb, L. A. (2018). Enhancing the accuracy of next-generation sequencing for detecting rare and subclonal mutations. *Nat. Rev. Genet.* 19 (5), 269–285. doi:10.1038/nrg.2017.117
- Sarwar, G., Sharif, I., Younas, A., Nazir, A., and Farooq, J. (2022). Historical work to combat with cotton leaf curl disease (clud) in Pakistan. *J. Agric. Res.* 60 (3), 03681157.
- Sattar, M. N., Kvarnheden, A., Saeed, M., and Briddon, R. W. (2013). Cotton leaf curl disease—an emerging threat to cotton production worldwide. *J. General Virology* 94 (4), 695–710. doi:10.1099/vir.0.049627-0
- Saunders, K., Norman, A., Gucciardo, S., and Stanley, J. (2004). The DNA beta satellite component associated with ageratum yellow vein disease encodes an essential pathogenicity protein (betaC1). *Virology* 324 (1), 37–47. doi:10.1016/j.virol.2004.03.018
- Savic, N., Ringnalda, F. C., Lindsay, H., Berk, C., Bargsten, K., Li, Y., et al. (2018). Covalent linkage of the DNA repair template to the CRISPR-Cas9 nuclease enhances homology-directed repair. *elife* 7, e33761. doi:10.7554/eLife.33761
- Schoen, A., Wallace, S., Holbert, M. F., Brown-Guidera, G., Harrison, S., Murphy, P., et al. (2023). Reducing the generation time in winter wheat cultivars using speed breeding. *Crop Sci.* 63 (4), 2079–2090. doi:10.1002/csc2.20989
- Shafique, T., Shafique, J., Shafique, A., Shafique, S., Shahid, M., Uddin, N., et al. (2022). Crispr/cas9 in gossypium hirsutum (cotton) coker 312 for Clud cotton leaf curl virus disease resistance mediated by Agrobacterium. *J. Microbiol. Mol. Genet.* 3 (3), 81–96.
- Sharma, N., and Prasad, M. (2020). Silencing AC1 of Tomato leaf curl virus using artificial microRNA confers resistance to leaf curl disease in transgenic tomato. *Plant Cell Rep.* 39, 1565–1579. doi:10.1007/s00299-020-02584-2
- Sharma, P., Rishi, N., and Malathi, V. G. (2005). Molecular cloning of coat protein gene of an Indian cotton leaf curl virus (CLCuV-HS2) isolate and its phylogenetic relationship with others members of Geminiviridae. *Virus Genes* 30, 85–91. doi:10.1007/s11262-004-4585-x
- Shuli, F., Jarwar, A. H., Wang, X., Wang, L., and Ma, Q. (2018). Overview of the cotton in Pakistan and its future prospects. *Pak. J. Agric. Res.* 31 (4), 396. doi:10.17582/journal.pjar/2018/31.4.396.407
- Singh, A., Vivek, A., Gupta, K., Sharma, S., and Kumar, S. (2023). Long non-coding RNA and microRNA landscape of two major domesticated cotton species. *Comput. Struct. Biotechnol. J.* 21, 3032–3044.
- Singh, I., Kaur, R., Kumar, A., Singh, S., and Sharma, A. (2021). Differential expression of gut protein genes and population density of *Arsenophonus* contributes to sex-biased transmission of *Bemisia tabaci* vectored Cotton leaf curl virus. *Plos one* 16 (11), e0259374. doi:10.1371/journal.pone.0259374
- Sodha, D., Verma, S. K., Chhokar, V., and Paul, D. (2022). Cotton lead curl viral disease in American cotton (*G. Hirsutum*): genetic basis of resistance and role of genetic engineering tools in combating CLCUD. *Scientist* 1 (3), 5122–5137.
- Song, Q., Zhang, T., Stelly, D. M., and Chen, Z. J. (2017). Epigenomic and functional analyses reveal roles of epialleles in the loss of photoperiod sensitivity during domestication of allotetraploid cottons. *Genome Biol.* 18, 99–14. doi:10.1186/s13059-017-1229-8
- Sony, S. K., Kaul, T., Motelb, K. F. A., Thangaraj, A., Bharti, J., Kaul, R., et al. (2023). CRISPR/Cas9-mediated homology donor repair base editing confers glyphosate resistance to rice (*Oryza sativa* L.). *Front. Plant Sci.* 14, 1122926. doi:10.3389/fpls.2023.1122926

- Sun, F., Ma, J., Wang, P., and Yang, Y. (2022). Genome-wide identification of the SAMS gene family in upland cotton (*Gossypium hirsutum* L.) and expression analysis in drought stress treatments. *Genes* 13 (5), 860. doi:10.3390/genes13050860
- Suthar, T., Gupta, N., Pathak, D., Sharma, S., and Rathore, P. (2021). *Morpho-anatomical characterization of interspecific derivatives of Gossypium hirsutum L. × G. armourianum* Kearney cross for whitefly tolerance. Springer: Phytoparasitica, 1–19.
- Tang, Y., Zhang, Z., Yang, Z., and Wu, J. (2023). CRISPR/Cas9 and Agrobacterium tumefaciens virulence proteins synergistically increase efficiency of precise genome editing via homology directed repair in plants. *J. Exp. Bot.*, erad096.
- Tao, X. Y., Xue, X. Y., Huang, Y. P., Chen, X. Y., and Mao, Y. B. (2012). Gossypol-enhanced P450 gene pool contributes to cotton bollworm tolerance to a pyrethroid insecticide. *Mol. Ecol.* 21 (17), 4371–4385. doi:10.1111/j.1365-294X.2012.05548.x
- Tatineni, S., and Hein, G. L. (2021). High Plains wheat mosaic virus: an enigmatic disease of wheat and corn causing the High Plains disease. *Mol. Plant Pathol.* 22 (10), 1167–1179.
- Thomas, J., Kim, H. R., Rahmatallah, Y., Wiggins, G., Yang, Q., Singh, R., et al. (2019). RNA-seq reveals differentially expressed genes in rice (*Oryza sativa*) roots during interactions with plant-growth promoting bacteria, *Azospirillum brasilense*. *PLoS One* 14 (5), e0217309. doi:10.1371/journal.pone.0217309
- Tian, S., Wang, X., Li, P., Wang, H., Ji, H., Xie, J., et al. (2016). Plant aquaporin AtPIP1; 4 links apoplastic H₂O₂ induction to disease immunity pathways. *Plant physiol.* 171 (3), 1635–1650. doi:10.1104/pp.15.01237
- Timnaz, S., and Batley, J. (2019). DNA methylation: toward crop disease resistance improvement. *Trends Plant Sci.* 24 (12), 1137–1150. doi:10.1016/j.tplants.2019.08.007
- Uniyal, A. P., Yadav, S. K., and Kumar, V. (2019). The CRISPR–Cas9, genome editing approach: a promising tool for drafting defense strategy against begomoviruses including cotton leaf curl viruses. *J. Plant Biochem. Biotechnol.* 28 (2), 121–132. doi:10.1007/s13562-019-00491-6
- Van Bel, M., Silvestri, F., Weitz, E. M., Kreft, L., Botzki, A., Coppens, F., et al. (2022). PLAZA 5.0: extending the scope and power of comparative and functional genomics in plants. *Nucleic Acids Res.* 50 (D1), D1468–D1474. doi:10.1093/nar/gkab1024
- Varma, A., and Malathi, V. (2003). Emerging geminivirus problems: a serious threat to crop production. *Ann. Appl. Biol.* 142 (2), 145–164. doi:10.1111/j.1744-7348.2003.tb00240.x
- Verchot, J. (2012). Cellular chaperones and folding enzymes are vital contributors to membrane bound replication and movement complexes during plant RNA virus infection. *Front. plant Sci.* 3, 275. doi:10.3389/fpls.2012.00275
- Vij, S., Pathak, D., Rathore, P., Kumar, H., Sekhon, P. S., Bhatia, D., et al. (2022). Molecular mapping of CLCuD resistance introgressed from synthetic cotton polyploid in upland cotton. *J. Genet.* 101 (1), 25. doi:10.1007/s12041-022-01365-y
- Wang, F., Wang, C., Yan, Y., Jia, H., and Guo, X. (2016). Overexpression of cotton GhMPK11 decreases disease resistance through the gibberellin signaling pathway in transgenic *Nicotiana benthamiana*. *Front. plant Sci.* 7, 689. doi:10.3389/fpls.2016.00689
- Wang, G., Xu, Z., Wang, F., Huang, Y., Xin, Y., Liang, S., et al. (2022). Development of an efficient and precise adenine base editor (ABE) with expanded target range in allotetraploid cotton (*Gossypium hirsutum*). *BMC Biol.* 20 (1), 45–15. doi:10.1186/s12915-022-01232-3
- Wang, J., Hu, Z., Zhao, T., Yang, Y., Chen, T., Yang, M., et al. (2015). Genome-wide analysis of bHLH transcription factor and involvement in the infection by yellow leaf curl virus in tomato (*Solanum lycopersicum*). *BMC genomics* 16, 39–14. doi:10.1186/s12864-015-1249-2
- Wang, K., Wang, Z., Li, F., Ye, W., Wang, J., Song, G., et al. (2012). The draft genome of a diploid cotton *Gossypium raimondii*. *Nat. Genet.* 44 (10), 1098–1103. doi:10.1038/ng.2371
- Wang, W., Niu, S., Dai, Y., Wang, M., Li, Y., Yang, W., et al. (2019). The Zea mays mutants opaque2 and opaque 16 disclose lysine change in waxy maize as revealed by RNA-Seq. *Sci. Rep.* 9 (1), 12265. doi:10.1038/s41598-019-48478-6
- Wang, W., Xu, J., Fang, H., Li, Z., and Li, M. (2020). Advances and challenges in medicinal plant breeding. *Plant Sci.* 298, 110573. doi:10.1016/j.plantsci.2020.110573
- Wang, Y., and Wu, W.-H. (2010). Plant sensing and signaling in response to K⁺-deficiency. *Mol. plant* 3 (2), 280–287. doi:10.1093/mp/ssq006
- Watson, A., Ghosh, S., Williams, M. J., Cuddy, W. S., Simmonds, J., Rey, M.-D., et al. (2018). Speed breeding is a powerful tool to accelerate crop research and breeding. *Nat. plants* 4 (1), 23–29. doi:10.1038/s41477-017-0083-8
- Wei, Y., Liu, Y., Ali, A. M., Xiao, R., Liang, C., Meng, Z., et al. (2022). Rich variant phenotype of *Gossypium hirsutum* L. saturated mutant library provides resources for cotton functional genomics and breeding. *Industrial Crops Prod.* 186, 115232. doi:10.1016/j.indcrop.2022.115232
- Weirather, J. L., de Cesare, M., Wang, Y., Piazza, P., Sebastiano, V., Wang, X.-J., et al. (2017). *Comprehensive comparison of pacific biosciences and Oxford nanopore technologies and their applications to transcriptome analysis*. 1000Research 6.
- Wen, J., Jiang, F., Weng, Y., Sun, M., Shi, X., Zhou, Y., et al. (2019). Identification of heat-tolerance QTLs and high-temperature stress-responsive genes through conventional QTL mapping, QTL-seq and RNA-seq in tomato. *BMC plant Biol.* 19, 1–17. doi:10.1186/s12870-019-2008-3
- Winter, P., and Kahl, G. (1995). Molecular marker technologies for plant improvement. *World J. Microbiol. Biotechnol.* 11, 438–448. doi:10.1007/BF00364619
- Wolter, F., Klemm, J., and Puchta, H. (2018). Efficient in planta gene targeting in *Arabidopsis* using egg cell-specific expression of the Cas9 nuclease of *Staphylococcus aureus*. *plant J.* 94 (4), 735–746. doi:10.1111/tpj.13893
- Xia, X. C., Hu, Q. Q., Li, W., Chen, Y., Han, L. H., Tao, M., et al. (2018). Cotton (*Gossypium hirsutum*) JAZ3 and SLR1 function in jasmonate and gibberellin mediated epidermal cell differentiation and elongation. *Plant Cell, Tissue Organ Cult. (PCTOC)* 133, 249–262. doi:10.1007/s11240-018-1378-9
- Xiong, X. P., Sun, S. C., Zhang, X. Y., Li, Y. J., Liu, F., Zhu, Q. H., et al. (2020). GhWRKY70D13 regulates resistance to *Verticillium dahliae* in cotton through the ethylene and jasmonic acid signaling pathways. *Front. Plant Sci.* 11, 69. doi:10.3389/fpls.2020.00069
- Xu, F., Chen, S., Zhou, S., Yue, C., Yang, X., Zhang, X., et al. (2022). Genome-wide association, RNA-seq and iTRAQ analyses identify candidate genes controlling radicle length of wheat. *Front. Plant Sci.* 13, 939544. doi:10.3389/fpls.2022.939544
- Xu, Y., Fu, S., Tao, X., and Zhou, X. (2021). Rice stripe virus: exploring molecular weapons in the arsenal of a negative-sense RNA virus. *Annu. Rev. Phytopathology* 59, 351–371. doi:10.1146/annurev-phyto-020620-113020
- Yadav, N., Yadav, D. K., Yadav, S., and Khurana, S. P. (2021). *Viral movement-cellular protein interaction*. *Plant Virus-Host Interaction*. Elsevier, 59–109.
- Yang, J., Zhang, H., Chen, H., Sun, Z., Ke, H., Wang, G., et al. (2023). Genome-wide association study reveals novel SNPs and genes in *Gossypium hirsutum* underlying Aphis gossypii resistance. *Theor. Appl. Genet.* 136 (8), 171. doi:10.1007/s00122-023-04415-w
- Yaqoob, S., Fatima, N., Khan, S., Ali, Q., Hafeez, M., and Malik, A. (2020). Begomoviruses and betasatellites associated with CLCuD. *Biol. Clin. Sci. Res. J.* 2020 (1). doi:10.54112/bcsrj.v2020i1.2
- Yazdani-Khameneh, S., Aboutorabi, S., Shoori, M., Aghazadeh, A., Jahanshahi, P., Golnaraghi, A., et al. (2016). Natural occurrence of tomato leaf curl New Delhi virus in Iranian cucurbit crops. *plant pathology J.* 32 (3), 201–208. doi:10.5423/PPJ.OA.10.2015.0210
- Yu, H., Lin, T., Meng, X., Du, H., Zhang, J., Liu, G., et al. (2021). A route to *de novo* domestication of wild allotetraploid rice. *Cell* 184 (5), 1156–1170. doi:10.1016/j.cell.2021.01.013
- Zargar, S. M., Raatz, B., Sonah, H., Bhat, J. A., Dar, Z. A., Agrawal, G. K., et al. (2015). Recent advances in molecular marker techniques: insight into QTL mapping, GWAS and genomic selection in plants. *J. crop Sci. Biotechnol.* 18 (5), 293–308. doi:10.1007/s12892-015-0037-5
- Zerbini, F. M., Briddon, R. W., Idris, A., Martin, D. P., Moriones, E., Navas-Castillo, J., et al. (2017). ICTV virus taxonomy profile: Geminiviridae. *J. general virology* 98 (2), 131–133. doi:10.1099/jgv.0.000738
- Zhang, H., Zhang, J., Xu, Q., Wang, D., Di, H., Huang, J., et al. (2020). Identification of candidate tolerance genes to low-temperature during maize germination by GWAS and RNA-seq approaches. *BMC plant Biol.* 20, 1–17. doi:10.1186/s12870-020-02543-9
- Zhang, L., Yu, Y., Zhang, M., Rong, K., Wu, Y., Zhang, M., et al. (2023). Genome-wide identification of xylan glucuronosyltransferase family in cotton and function characterization of GhGUX5 in regulating *Verticillium* wilt resistance. *Int. J. Biol. Macromol.* 245, 124795. doi:10.1016/j.ijbiomac.2023.124795
- Zhang, S., Tian, Z., Li, H., Guo, Y., Zhang, Y., Roberts, J. A., et al. (2019a). Genome-wide analysis and characterization of F-box gene family in *Gossypium hirsutum* L. *BMC genomics* 20, 1–16. doi:10.1186/s12864-019-6280-2
- Zhang, Y., Malzahn, A. A., Sretenovic, S., and Qi, Y. (2019b). The emerging and uncultivated potential of CRISPR technology in plant science. *Nat. Plants* 5 (8), 778–794. doi:10.1038/s41477-019-0461-5
- Zhong, Y., Cheng, C., Jiang, B., Jiang, N., Zhang, Y., Hu, M., et al. (2016). Digital gene expression analysis of Ponkan Mandarin (*Citrus reticulata* Blanco) in response to Asia citrus psyllid-vectored Huanglongbing infection. *Int. J. Mol. Sci.* 17 (7), 1063. doi:10.3390/ijms17071063
- Zhou, L., He, H., Liu, R., Han, Q., Shou, H., and Liu, B. (2014). Overexpression of GmAKT2potassium channel enhances resistance to soybean mosaic virus. *BMC plant Biol.* 14 (1), 1–11. doi:10.1186/1471-2229-14-154
- Zhou, X. (2013). Advances in understanding begomovirus satellites. *Annu. Rev. phytopathology* 51, 357–381. doi:10.1146/annurev-phyto-082712-102234
- Zhou, X., Xie, Y., Tao, X., Zhang, Z., Li, Z., and Fauquet, C. M. (2003). Characterization of DNAbeta associated with begomoviruses in China and evidence for co-evolution with their cognate viral DNA-A. *J. General virology* 84 (1), 237–247. doi:10.1099/vir.0.18608-0
- Zsögön, A., Čermák, T., Naves, E. R., Notini, M. M., Edel, K. H., Weinl, S., et al. (2018). *De novo* domestication of wild tomato using genome editing. *Nat. Biotechnol.* 36 (12), 1211–1216. doi:10.1038/nbt.4272

Frontiers in Genetics

Highlights genetic and genomic inquiry relating to all domains of life

The most cited genetics and heredity journal, which advances our understanding of genes from humans to plants and other model organisms. It highlights developments in the function and variability of the genome, and the use of genomic tools.

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

