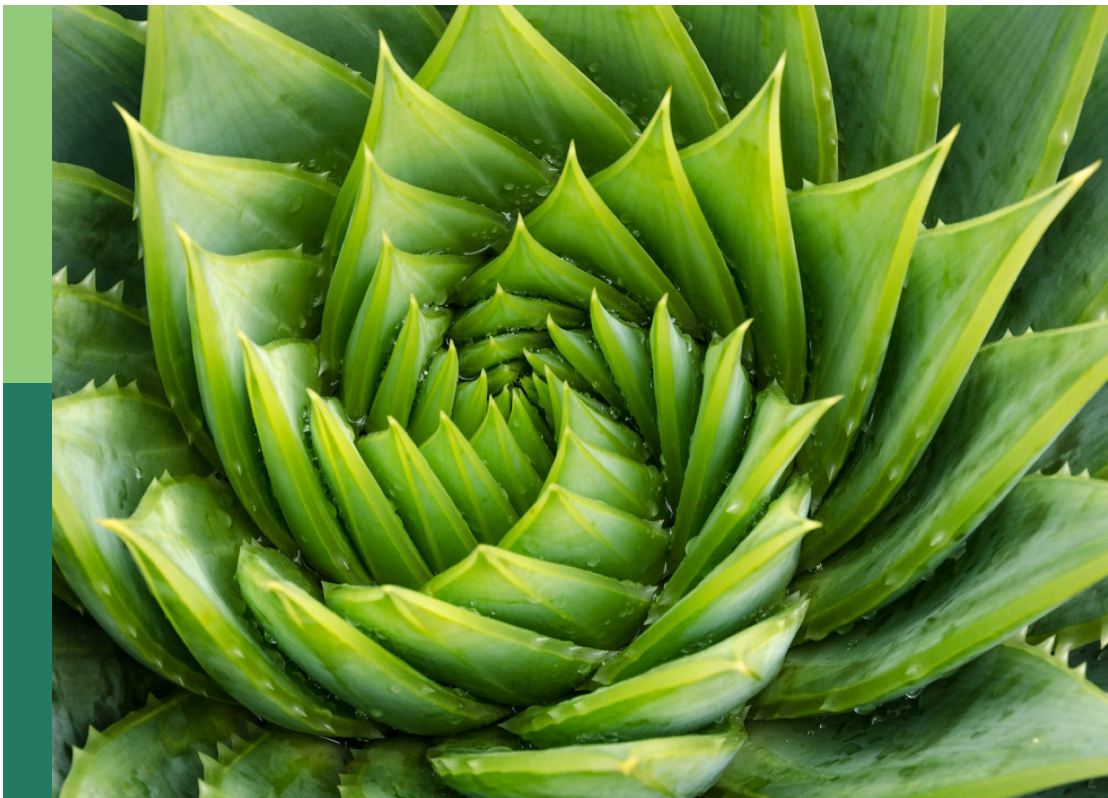
frontiers | Research Topics

Plant biotechnology and genetics for sustainable agriculture and global food security

Edited by
Galal Bakr Anis, Yinglong Chen and Baohong Zhang

Published in
Frontiers in Plant Science



ISSN 1664-8714
ISBN 978-2-8325-5420-3
DOI 10.3389/978-2-8325-5420-3

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

Plant biotechnology and genetics for sustainable agriculture and global food security

Topic editors

Galal Bakr Anis — Field Crops Research Institute, Agricultural Research Center, Egypt

Yinglong Chen — University of Western Australia, Australia

Baohong Zhang — East Carolina University, United States

Citation

Anis, G. B., Chen, Y., Zhang, B., eds. (2024). *Plant biotechnology and genetics for sustainable agriculture and global food security*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-5420-3

Table of contents

frontiers | Frontiers in Plant Science

TYPE Editorial
PUBLISHED 30 August 2024
DOI 10.3389/fpls.2024.1479632

OPEN ACCESS

EDITED AND REVIEWED BY
James Lloyd,
Stellenbosch University, South Africa

*CORRESPONDENCE
Yinglong Chen
yinglong.chen@uwa.edu.au

RECEIVED 12 August 2024
ACCEPTED 19 August 2024
PUBLISHED 30 August 2024

CITATION
Chen Y, Zhang B and Annis GB (2024) Editorial: Plant biotechnology and genetics for sustainable agriculture and global food security.
Front. Plant Sci. 15:1479632.
doi: 10.3389/fpls.2024.1479632

Editorial: Plant biotechnology and genetics for sustainable agriculture and global food security

Yinglong Chen[1]*, Baohong Zhang[2] and Galal Bakr Annis[3]

[1]School of Agriculture and Environment, and UWA Institute of Agriculture, The University of Western Australia, Perth, WA, Australia, [2]Department of Biology, East Carolina University, Greenville, NC, United States, [3]Field Crops Research Institute, Agricultural Research Center, Giza, Egypt

KEYWORDS

plant biotechnology and breeding, genetics, genome-wide association, next generation sequencing, CRISPR, omics, agriculture sustainability, genotyping (GT)

Editorial on the Research Topic
Plant biotechnology and genetics for sustainable agriculture and global food security

The global landscape of food security is increasingly precarious, exacerbated by multifaceted challenges ranging from conflict-driven crises to the compounding impacts of climate change and economic instability. According to the World Food Programme, people facing acute hunger has nearly doubled since 2019, from 135 million to 258 million, a number that could potentially double again due to the disruptive effects of the COVID-19 pandemic, placing an additional 130 million at risk (Ramanujam et al., 2024). Moreover, pervasive food insecurity not only jeopardizes public health with its links to malnutrition-related illnesses and premature deaths but also underscores a critical need for enhanced agricultural productivity.

Since the inception of the Green Revolution, strides in crop productivity have been significant but insufficient. Annual yield increases for major crops currently hover between 0.8% to 1.2%, a pace that falls short of what is required to sustainably feed a burgeoning global population (Ahmad, 2023). Classical breeding techniques have historically played a pivotal role in enhancing crop varieties and ensuring food security. However, their limitations in meeting contemporary global demands have spurred a drive towards novel plant breeding techniques (NPBTs) and other advanced agricultural methodologies.

In recent years, breakthroughs such as genome-wide association studies (GWAS), Next Generation Sequencing (NGS), and genome editing (GE) have emerged as transformative tools in crop improvement efforts (Mangrauthia et al., 2024). These technologies offer unprecedented access to vast gene pools, enabling targeted enhancements in traits essential for crop resilience, yield, and nutritional quality. Versatile techniques, such as Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/CRISPR-associated protein (CRISPR/Cas) genome editing, marker-assisted selection, and Quantitative Trait loci (QTL) mapping exemplify the precision and potential of NPBTs to fortify crops against various environmental biotic and abiotic stresses while optimizing productivity.

This Research Topic explores cutting-edge research in novel plant breeding strategies, with a focus on bolstering crop tolerance to abiotic and biotic stresses. It elucidates the mechanistic underpinnings of gene function and regulation with the advancements in CRISPR/Cas-based editing techniques and integrative multi-omics analyses.

Advanced plant biotechnology and genetics approaches

Traditional methods in quantitative genetics, including GWAS, genomic selection, QTL mapping, have historically formed the backbone of crop breeding programs (Maldonado et al., 2020; Thudi et al., 2021). Compared to Bayesian approaches, the efficacy of deep learning models for predicting flowering-related traits in a tropical maize panel from Brazil using single nucleotide polymorphisms (SNPs) data across multiple environments was evaluated in a study of Mora-Poblete et al. Multi-trait models showed a 14.4% increase in prediction accuracy over single-trait models, while multi-environment schemes improved accuracy by 6.4%. Deep learning consistently outperformed Bayesian methods in both scenarios. Genome-wide association analysis identified significant marker-trait associations, highlighting the potential of deep learning in enhancing genomic selection for improving maize breeding programs focused on flowering traits.

The challenge of maintaining crop health in the face of evolving pathogens is outlined in a review of Yıldırım et al., which emphasizes the limitations of traditional breeding methods against emerging diseases exacerbated by climate change. It underscores the promising role of genome editing in enhancing crop resilience by targeting defense-related genes, thereby advancing sustainable agriculture and bolstering food security through durable disease resistance strategies. The study advocates for genome editing as a pivotal tool in realizing the concept of "healthy plants" capable of withstanding diverse biotic stresses.

Using k-mer-based sequence comparison and deep learning-based variant calling approaches, Ruperao et al. investigated genetic and morphological variation in sorghum race populations. The analysis of 272 sorghum accessions revealed 1.7 million high-quality SNPs and identified 2,370 genes associated with selection signatures, including 179 selective sweep regions across 10 chromosomes. These findings highlight the genetic basis of domestication traits such as biomass and plant height, with implications for future plant breeding programs.

Through computational analysis and expression profiling technologies, Tan et al. identified 37 potassium (K^+) transport-related genes (PTGs) in mango (*Mangifera indica*), including 22 K^+ transporters and 15 K+ channels, essential for plant growth, development, and drought tolerance. Phylogenetic and promoter analyses revealed strong kinship with other plants and identified cis-elements related to various biological processes. RNA-seq expression profiling highlighted specific PTGs upregulated in roots and leaves and during different growth stages, providing foundational knowledge for understanding K^+ transport in mango and aiding in the functional characterization of K^+ genes in tropical fruits.

In a review of Das et al., the authors highlight that plants are a crucial source of specialized metabolites, offering physiological benefits and evolutionary advantages, especially in defence mechanisms. Advances in transgenic techniques like gene silencing and overexpression have boosted the production of these compounds, reducing costs and enhancing nutritional value. The use of CRISPR/Cas-based gene editing is now pivotal in improving the yield of specific metabolites in medicinal plants, offering significant advancements in metabolic engineering for future applications.

Plant biotechnology and genetics reveal plant responses to environmental stress

A study reported in Geldhof et al. investigates how tomato plants adapt to waterlogging through downward leaf bending, a trait influenced by complex genetic and environmental interactions. Using a GWAS of 54 tomato accessions, researchers identified candidate genes linked to plant survival under waterlogged conditions. The study highlighted genes potentially involved in metabolic adjustments and leaf angle dynamics, suggesting their roles in facilitating plant resilience and recovery from waterlogging stress.

In the review of Devi et al. the deficiency of methionine in maize, a critical amino acid essential for animal and human nutrition, is addressed. This review explores biofortification strategies, emphasizing the role of zein proteins, particularly δ-zein, in methionine deposition within maize kernels. The study consolidates various approaches including natural mutant selection, genetic engineering techniques, and meta-QTL studies to enhance methionine content in maize, offering insights for future research aimed at developing high-methionine maize varieties to improve nutritional outcomes sustainably.

Shang et al. identified and characterized 82 class III peroxidase (PRX) proteins in sugarcane, revealing their role in lignification, cell elongation, seed germination, and stress responses. Phylogenetic and promoter analyses showed the evolutionary history and regulatory elements of ShPRX genes, highlighting their response to ABA, MeJA, light, anaerobic conditions, and drought. Differential expression and qRT-PCR analyses demonstrated that these genes are specifically induced by SCMV, Cd, and salt, providing insights into their structure, evolution, and potential for developing stress-resistant sugarcane varieties and phytoremediation strategies.

Ma et al. summarised the multifaceted research on pennycress (*Thlaspi arvense* L.), an emerging oil crop within the Brassicaceae family. The review highlighted several key research areas including its utility as a model plant akin to *Arabidopsis thaliana*, advancements in oil and protein extraction technologies, metabolomics-based seed composition analysis, germplasm development, ecological impacts, abiotic stress responses, and

strategies for optimizing fatty acid extraction. Future research directions proposed include assembling the pennycress genome, refining biodiesel extraction processes, investigating molecular mechanisms of fatty acid synthesis, and elucidating the roles of pivotal genes under diverse environmental stresses.

Liu et al. investigated how different host plants influence the biological traits, nutritional metabolism, and *Buchnera* symbiont dynamics in the pea aphid (*Acyrthosiphon pisum*). It reveals significant variations in *Buchnera* titers across host plants, with higher levels observed on *T. pratense* and *M. officinalis*. Pea aphids on broad bean (*Vicia faba*) exhibited increased levels of soluble sugar, glycogen, and total energy, correlating with enhanced fecundity and weight. These findings underscore the complex interplay between host plant species, symbiotic interactions, and nutritional outcomes in aphid populations, offering insights into broader implications for insect development and management strategies.

Travadi et al. validated rbcL and ITS2 metabarcoding primers for detecting plant species in herbal products, using mock controls of medicinal plant DNA and biomass pools. This study demonstrates high detection efficiencies of 86.7% and 71.7% for rbcL, and 82.2% and 69.4% for ITS2, in DNA and biomass pools respectively. Combining these metabarcodes achieved a cumulative detection efficiency of 100% in DNA pools and 90% in biomass pools. Overall, the study underscores the potential of multilocus metabarcoding as a robust tool for identifying labelled and unlabelled plant species in herbal formulations, thereby enhancing pharmacovigilance in the global herbal medicine market.

Conclusions

Recent advancements in plant biotechnology and genetics underscore a transformative era in crop improvement and sustainable agriculture. Traditional methods such as GWAS, genomic selection, and QTL mapping continue to play pivotal roles in enhancing crop breeding programs, facilitating targeted trait improvements across diverse environments. The integration of deep learning models has shown promising results, thereby enriching the precision and efficiency of genomic selection. Furthermore, innovations in genome editing technologies offer new avenues for enhancing crop resilience against evolving pathogens and environmental stresses, contributing to global food security. Studies in this Research Topic highlight the evolving landscape of plant biotechnology, emphasizing the critical intersection of genetics, computational analysis, and biotechnological tools in shaping the future of agricultural sustainability and productivity.

Author contributions

YC: Writing – review & editing, Writing – original draft, Conceptualization. BZ: Writing – review & editing. GB: Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. YC was supported by Australian Research Council (FT210100902).

Conflict of interest

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

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

Publisher's note

References

Ahmad, M. (2023). Plant breeding advancements with "CRISPR-Cas" genome editing technologies will assist future food security. *Front. Plant Sci.* 14. doi: 10.3389/fpls.2023.1133036

Maldonado, C., Mora-Poblete, F., Contreras-Soto, R. I., Ahmar, S., Chen, J.-T., do Amaral Júnior, A. T., et al. (2020). Genome-wide prediction of complex traits in two outcrossing plant species through deep learning and bayesian regularized neural network. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.593897

Mangrauthia, S. K., Molla, K. A., Sundaram, R. M., Chinnusamy, V., and Bansal, K. C. (2024). "Genomics and Genome Editing for Crop Improvement," in *Transformation of Agri-Food Systems* (Springer Nature Singapore, Singapore), 297–322.

Ramanujam, N., Caivano, N., Agnello, A., and Neranjan, K. (2024). COVID-19, the SDG agenda, and implementation paralysis: cash transfer programs to the rescue? *McGill J. Sustain Dev. Law* 20, 1–27. Available at: https://ia-forum.org/Files/RRFUWN.pdf.

Thudi, M., Chen, Y., Pang, J., Kalavikatte, D., Bajaj, P., Roorkiwal, M., et al. (2021). Novel genes and genetic loci associated with root morphological traits, phosphorus-acquisition efficiency and phosphorus-use efficiency in chickpea. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.636973

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 23 January 2023
DOI 10.3389/fpls.2022.1102201

OPEN ACCESS

EDITED BY
Galal Bakr Anis,
Agricultural Research Center, Egypt

REVIEWED BY
Saira Bano,
University of Sargodha, Pakistan
Wei Li,
Agricultural Genomics Institute at Shenzhen (CAAS), China

*CORRESPONDENCE
Farrukh Azeem
azeuaf@hotmail.com

SPECIALTY SECTION
This article was submitted to Plant Biotechnology, a section of the journal Frontiers in Plant Science

RECEIVED 18 November 2022
ACCEPTED 21 December 2022
PUBLISHED 23 January 2023

CITATION
Tan L, Waqas M, Rehman A, Rashid MAR, Fiaz S, Manzoor H and Azeem F (2023) Computational analysis and expression profiling of potassium transport-related gene families in mango (*Mangifera indica*) indicate their role in stress response and fruit development. *Front. Plant Sci.* 13:1102201. doi: 10.3389/fpls.2022.1102201

Computational analysis and expression profiling of potassium transport-related gene families in mango (*Mangifera indica*) indicate their role in stress response and fruit development

Lin Tan[1], Muhammad Waqas[2], Abdul Rehman[2], Muhammad Abdul Rehman Rashid[2], Sajid Fiaz[3], Hamid Manzoor[4] and Farrukh Azeem[2]*

[1]Haikou Experimental Station, Chinese Academy of Tropical Agricultural Sciences, Haikou, Hainan, China, [2]Department of Bioinformatics and Biotechnology, Government College University, Faisalabad, Pakistan, [3]Department of Plant Breeding and Genetics, The University of Haripur, Haripur, Pakistan, [4]Institute of Molecular Biology and Biotechnology, Bahauddin Zakariya University, Multan, Pakistan

Mango (*Mangifera indica*) fruit is known for its taste, health benefits, and drought tolerance. Potassium (K^+) is one of the most abundant ions in a plant cell. It is important for various biological functions related to plant growth, development, and flowering/fruiting. It significantly contributes to fruit yield, quality, and drought tolerance in plants. However, molecular mechanisms comprising K^+ transport in mango are least known. In the present study, 37 members of K^+ transport-related genes (PTGs) were identified in mango, which include 22 K^+ transporters (16 HAKs, 1 HKT, and 6 KEAs) and 15 K^+ channels (6 TPKs and 8 Shakers). All PTGs were predicted to be expressed at the plasma membrane and possess characteristic motifs and domains. Phylogenetic analysis identified a strong kinship of PTGs among *Oryza sativa*, *Arabidopsis thaliana*, *Cicer arietinum*, *Malus domestica*, and *M. indica*. The promoter analysis identified 60 types of *cis*-elements related to various biological processes. RNA-seq-based expression profiling identified that *MiTPK1.2*, *MiHAK1*, *MiHAK2.1*, *HAK6.1*, and *MiAKT1.1* were most upregulated in roots and that *MiKEA2*, *MiAKT2*, and *MiAKT1* were upregulated in leaves. Moreover, *MiAKT6*, *MiHAK1.1*, *MiKAT2*, *MiKAT2.1*, *MiHKT1*, *MiTPK1.1*, *MiHAK7*, and *MiHAK12* were highly expressed during the five growth stages of mango fruit. The current study is the first comprehensive report on K^+ transport system in tropical fruits. Therefore, it will provide the foundation knowledge for the functional characterization of K^+ genes in mango and related plants.

KEYWORDS

mango, potassium, channel, transporter, fruit development, drought stress

1 Introduction

Mango (*Mangifera indica*) is a widespread, evergreen, and one of the most dominant tropical fruits worldwide, being the sixth most cultivated fruit after bananas (*Musa acuminata*), watermelons (*Citrullus lanatus*), apples (*Malus domestica*), oranges (*Citrus sinensis*), and grapes (*Vitis vinifera*) (Ngamchuachit et al., 2015). Its bright color, distinctive quality, unique taste, and nutritional value have promoted its consumption. The mango fruit trade continues to grow and develop in the food service and dime markets (Celestin, 2019). Mango can be used as a pudding, fresh juice, and extracted products; refined into jam; or used as a jelly bean. With extensive cultivation of this fruit, India is known as a major producer with 25 million tonnes in 2020 (FAOSTAT, 2022), as compared to other countries like Pakistan, China, Indonesia, Malawi, and Mexico (Galán Saúco, 2002). Mango trees can withstand drought conditions, but these may affect the overall quality of the fruit (Cleveland, 2012). Since peak fruit development occurs in the dry season, water requirement is critical (Singh and Kushwaha, 2006; Snowden et al., 2014). The standard ranking of export mangoes (flavor, size, shape, and color) can contribute an additional 30%–50% to the payout (Laurio, 2021). Farmers are encouraged to increase the irrigation for this crop because of its tremendous impact on profitable yield. However, water scarcity and the enormous expenditure of energy required to raise irrigation water during the peak season have threatened fruit production and cultivation (Zikki, 2020). For this reason, solid agronomic and irrigation practices are applied at the farm level to help plants survive drought stress. However, investigations into mango drought stress tolerance are relatively scarce.

K^+ acquisition is one of the most important issues covering organic agriculture, and it is because most organic sources of K^+ are poorly soluble, limiting plant growth (Council, 1993). Due to its role in protein synthesis, ionic stability, photosynthesis, stress tolerance, translocation of photosynthates, and the initiation of several plant enzymes, K^+ is an essential component for plant maturation and final production (Hossain et al., 2021; Ul Hassan et al., 2021). K^+ plays a unique role in the generation of starch, blooming, and fruit output (Farooq et al., 2014). However, increased K^+ rates in plants can result in an imbalance in magnesium and calcium nutrition (Nahar et al., 2015). For plant growth, the main sources are the utilization of chemical K^+ fertilizer and the disintegration of K^+-minerals (Etesami et al., 2017; Bahadur et al., 2019). The application of appropriate K^+ fertilizer to mango trees increases fruit yield and value. Nevertheless, unconstrained fertilizer usage can result in financial loss to farmers, and excessive use of non-renewable resources is raising concerns about large-scale sustainable development (Bahadur et al., 2019; Wang et al., 2022). Foliar application of K^+ is also practiced to improve fruit yield. Therefore, to amend the K^+ utilization for mango, there is a need to acknowledge K^+ transport mechanisms.

Different proteins in plants that control cellular K^+ absorption and distribution include both channels and transporters. Voltage-dependent channel proteins mediating K^+ transport include shaker-like channels, voltage-independent tandem-pore K^+ (TPK) channels, and two-pore channels (TPCs) (Voelker et al., 2010). Five subgroups are used to classify the Shaker family: weak inward rectifying channels, inward rectifying channels with KAT-like characteristics, outward rectifying channels, inward rectifying channels with AKT-like characteristics, and the silent/regulatory subunit (González et al., 2015). Moreover, the carrier-like KUP/HAK/KT family (Azeem et al., 2022), HKT uniporters and symporters (Sahi et al., 2022), and KEA antiporters (Wang et al., 2017) also manifested their involvement in this mechanism. The KUP/HAK/KT family may regulate K^+ with great affinity (Ye et al., 2013). The assimilation and homeostasis of K^+ and sodium ions depend on the HKT proteins. In plants, there are two kinds of HKTs, namely, HKT1-like and HKT 2-like. The Na^+ uniporters make up class I HKTs, whereas Na^+ and K^+ symporters make up class II HKTs (Riedelsberger et al., 2021). The KEA proteins are probably similar to bacterial KefC K^+/H^+ antiporters (Waters et al., 2013; Assaha et al., 2015). Most KEAs are shown to control the pH value of stroma and thylakoids at the chloroplast membrane (Sun et al., 2018). In endomembrane cells, KEA 4, 5, and 6 help to maintain pH and K^+ homeostasis in a balanced manner (Sze and Chanroj, 2018; Zhu et al., 2018). Pore domains (PDs) are used to calculate the K^+ channels, which are heterodimeric proteins with trans-membrane sections. Functional multimeric proteins are associated with four PDs to make a navigation pathway of channels. P domain of the K^+ channels contains a highly conserved motif, i.e., "GYGD/E". In *Arabidopsis thaliana*, 15 K^+-selective channels, including one K^+ inward rectifier (Kir-like), five tandem-pore K^+ channels (TPK), and nine voltage-gated ion channels, are grouped into three families based on their physiography. Additionally, K^+ transporters are divided into three families—the Trk/HKT family of high-affinity K^+ transporters, the KEA (K^+/H^+ antiprotons) family of K^+ efflux antiporters, and the KUP/HAK/KT family of K^+ uptake permeases—which collectively have 13 members (1 member) (Aranda-Sicilia et al., 2012).

The current study was designed to identify key protein families involved in K^+ transport. For this purpose, a combination of bioinformatics approaches was used to characterize potential PTGs, and expression profiling was conducted using next-generation sequencing (NGS) data available at the National Center for Biotechnology Information (NCBI).

2 Material and methods

2.1 Identification and sequence analysis of K^+ transport-related gene families in *M indica*

For the identification of PTG families in *M. indica*, PTG sequences from *A. thaliana*, *Cicer arietinum*, and *Cajanus cajan* were used to perform a BLAST search in the NCBI (https://www.ncbi.nlm.nih.gov) and Phytozome (https://phytozome-next.jgi.doe.gov/) databases (Johnson et al., 2008; Goodstein et al., 2012). An *E*-value $1e^{-10}$ was used as the cutoff value BLAST search. After retrieving all the sequencing, we utilized SMART and pfam tools to verify the accuracy of K^+ transporting gene family in *M. indica* and deleted the sequences that lacked the conserved domains. To enhance the accuracy of our analysis, we deleted genes having common start positions in the genome representing the same locus or splice variants.

2.2 Multiple sequence analysis and phylogenetic analysis of K^+ transporting gene family

The MEGA7 software with default parameters was used for multiple sequence alignment of PTG sequences of five different species, i.e., *A. thaliana*, *M. domestica*, *Oryza sativa*, *C. arietinum*, and *M. indica*. Moving forward after multiple sequence alignment (MSA) of 198 protein sequences of the PTG families, with the use of the IQ TREE web server (http://iqtree.cibiv.univie.ac.at/), a phylogenetic tree was built based upon the maximum likelihood method. The tree was designed and visualized by using a web-based tool ITOL (https://itol.embl.de/) (Schultz et al., 2000).

2.3 Physiochemical properties of K^+ transporting gene family members in *M. indica*

The physicochemical properties of PTG family members were predicted by using a web-based tool ProtParam (https://web.expasy.org/protparam/) with default options. Gene ID, chromosomal locations, number of nucleotides in genomic DNA and mRNA, and number of exons were identified by using the NCBI gene database (https://www.ncbi.nlm.nih.gov/gene). The subcellular localization of K^+ transporting proteins was predicted by using a web-based tool ProtCamp 9.0 (ProtComp—Predict the sub-cellular localization for Plant proteins (softberry.com), and then, these locations were also verified by another online tool, CELLO v.2.5 (http://cello.life.nctu.edu.tw/), by using protein sequences of K^+ transporting gene family (Garg et al., 2016; Mehanny et al., 2022).

2.4 Sequence analysis of PTG family members in *M. indica*

To predict the exons and introns of the K^+ transporting genes, genomic DNA and cDNA sequences of all the members of K^+ transporting genes of *M. indica* were retrieved from the NCBI database in FASTA format. Exons and introns in these genes were predicted using an online tool named Gene Structure Display Server (GSDS) 2.0 (http://gsds.gao-lab.org/). Motifs of the K^+ transporting gene family were identified by an online software called MEME (https://meme-suite.org/meme/tools/meme). To predict the conserved motifs, the following parameters were selected: one occurrence per sequence (oops), number of motifs of 10, motifs with at least a width of 10, and utmost width of 50. Motif logos were also made by MEME (Guo et al., 2007; Bailey et al., 2009).

2.5 Domain analysis of K^+ transporting genes in *M. indica*

To predict the conserved domains of the K^+ transporting gene family in *M. indica*, the NCBI-CDD database (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) was used in which all the sequences were searched against the Pfam database. These results were also verified by the SMART tool (http://smart.embl-heidelberg.de/). Then, conserved domains of the K^+ transporting gene family were visualized by the TB tool, using the results of the Pfam database.

2.6 *Cis*-regulatory elements of K^+ transporting gene family in *M. indica*

For determining, the *cis*-acting regulatory elements of the K^+ transporting genes in *M. indica*, an approximately 2,000-bp promoter region was retrieved for each gene from the NCBI gene page. After the promoter sequence was retrieved, the PlantCare database (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) was employed to predict the *cis*-acting regulatory elements in promoter regions of K^+ transporting genes. The PlantCare results were downloaded and opened in the Excel sheets that were used to anticipate the *cis*-acting regulatory elements in TBtools software (Lescot et al., 2002).

2.7 Chromosomal distribution of K^+ transporting genes

The gene location (initial and terminating position) of K^+ transporting genes present on the chromosomes was identified

by the NCBI gene database. The chromosome number for each member of the PTG family was also identified from the NCBI database. The chromosomal location was visualized by a Ritchie lab tool, Phenograms (http://visualization.ritchielab.org/phenograms/plot).

2.8 Gene expression profile of K^+ transporting genes using NGS data and qRT-PCR

For evaluation of the gene expression configuration of the K^+ transporting genes under abiotic stress (drought stress with bio-project), genome and gff files were downloaded from the NCBI-SRA [Home—SRA—NCBI (nih.gov)] database after the genome, and gff files were retrieved from the NCBI-genome database. The galaxy server [Galaxy (usegalaxy.org)] was utilized to obtain the fragments per kilobase of transcript per million mapped reads (FPKM) values, and the expression pattern was shown as a heat map using TB tools. Drought stress was applied to 1-year-old mango plants (Chaunsa) grown in soil-filled pots (one plant/pot) under an ambient environment (12-h photoperiod, 50%–60% relative humidity, 25° C ± 1°C day/night temperature). Drought stress was imposed by watering each pot (n = 6) with 1 L of 200 g/L PEG-6000. Leaves were collected 2, 4, and 7 days after the start of treatment. To validate the expression profiles of NGS data, real-time RT-qPCR was used. For this purpose, RNA was extracted from mango leaves using the TRIzol reagent and was quantified using NanoDrop 2000 (Thermo Fisher Scientific, Waltham, MA, USA). With the use of 1 µg of RNA and the Maxima H-minus First-Strand cDNA synthesis kit, the RNA was reverse transcribed, and cDNA was stored at −20° C. An iTaq Universal SYBR Green Super-Mix and a qRT-PCR detection equipment (CFX96 Touch RT PCR Detection System, Bio-Rad Labs, Hercules, CA, USA) were used to perform the qRT-PCR. Gene-specific primers were mapped using an online program "Oligo Calculator" at mcb.berkeley.edu/labs/krantz/tools/oligocalc.html (accessed 26 July 2022), and the specificity of primers was verified using NCBI-primer BLAST algorithm (https://www.ncbi.nlm.nih.gov/tools/primer-blast/) (retrieved on 26 July 2022). Actin (LOC123192663) was used as a reference gene (Yao et al., 2022). The $2^{-\Delta\Delta CT}$ method is used to calculate respective gene expression levels on the basis of three biological replicates (Livak and Schmittgen, 2001).

3 Results

3.1 Computational identification and characterization of K^+ transport-related genes

A total of 37 PTGs were identified in *M. indica* genome (Table 1). Physiochemical characteristics of K^+ transporter in *M. indica*, such as the total number of amino acids (aa), molecular weights (MW), aliphatic index, gravy, and hypothetical isoelectric points (pI), were identified using the web-based tool Expasy-protparam. Subcellular localization was also identified using cello 2.0. The length of protein ranged from 340aa to 1208aa. Molecular weight varied from 39.05 to 130.77 kDa, and the pI values start from 4.92 to 9.47 in which 16 of the proteins were acidic and 20 were basic. Interestingly, the MiHAK1.1 protein possessed a pI value of 7.01. GRAVY values of AKT and KAT-like proteins were negative (hydrophilic), while the rest of the proteins were hydrophobic, and subcellular localization showed that the location of these proteins is in the plasma membrane (Table 1).

3.2 Phylogenetic analysis of K^+ transport-related proteins in *M. indica*

To predict the functional properties as well as phylogenetic relationships of the K^+ transport-related genes, using protein sequences of PTGs of *O. sativa*, *A. thaliana*, *C. arietinum*, *M. domestica*, and *M. indica*, a phylogenetic tree was constructed. The members of the tree were divided into five groups: i) HAK, ii) HKT, iii) KEA, iv) TPK, and v) Shakers (AKT and KAT, and GORK and SKOR). HAK, HKT, and KEA are transporters, and TPK and Shakers are Channels. In phylogenetic analysis, some orthologous pairs (*MiHAK1*/*MdHAK1.2* and *MdHAK6.2*/*MiHAK6.1*), co-orthologous groups (*MdHAK1.1*/*CarHAK1*/*MiHAK1.1*, *MdHAK7.1*/*MdHAK7.2*/*MiHAK7*, and *MdHAK5.4*/*CarHAK5*/*MiHAK5*), and paralogous groups (*MiHAK8*/*MiHAK8.1*, *MiHAK2.1*/*MiHAK2*, and *MiHAK5.3*/*MiHAK5.1*/*MiHAK5.2*) of *M. indica* were identified in the HAK family among *A. thaliana*, *O. sativa*, and *M. domestica*. In the HKT family, one orthologous pair was identified in *A. thaliana* (*MiHKT1*/*ATHKT1*). In the KEA family, two orthologous pairs (*MiKEA6*/*CarKEA6* and *MiKEA4*/*OsKEA4*) and two co-orthologous groups (*MdKEA3.1*/*MdKEA3.2*/*MiKEA3* and *MdKEA5.1*/*CarKEA5*/*MiKEA5*) were identified in the genome of *M. indica*. In the AKT and KAT sub-families, paralogous (*MiAKT1*/*MiAKT1.1* and *MiKAT2*/*MiKAT2.1*) and orthologues groups (*AtAKT6*/*MiAKT6*, *MiAKT2*/*MdAKT2.1*, and *MdKAT3.1*/*MdKAT3.2*/*MiKAT3*) were identified in *M. indica*. In the GORK and SKOR sub-families, one orthologous pair (*MiSKOR1*/*CarGORK*) was identified; in the TPK family, some orthologous and co-orthologous groups were identified (*MdTPK4.1*/*MdTPK4.2*/*MiTPK4*, *MiTPK5*/*CarTPK5*, and *MiTPK1.2*/*MdTPK1.3*); some paralogous groups were also identified (*MiTPK6*/*MiTPK6.1* and *MiTPK1.1*/*MiTPK1*) in the *M. indica* genome (Figure 1).

3.3 Gene structure and motif detection of K^+ transport-related proteins in *M indica*

To acquire an understanding of the structural properties of PTGs, we explored the intron and exon architecture of these

TABLE 1 Physiochemical properties of K^+ transport-related genes in *Mangifera indica*.

Sr. no.	Accession no.	Gene names	Gene ID	Chr. No.	Exon	Protein length (aa)	Molecular weight (Da)	pI	Gravy	Aliphatic index	Subcellular localization
1	XP_044508080.1	*MiHKT1*	LOC123227371	10	3	498	56,848.8	9.47	0.298	104	Plasma membrane
2	XP_044473757.1	*MiHAK1.1*	LOC123202089	2	9	753	83,293.7	7.01	0.489	111	Plasma membrane
3	XP_044493735.1	*MiHAK6.1*	LOC123217052	5	8	781	87,042	8.59	0.336	108.92	Plasma membrane
4	XP_044466396.1	*MiHAK6*	LOC123196431	14	8	776	86,449.4	8.15	0.377	109.11	Plasma membrane
5	XP_044508633.1	*MiHAK2.1*	LOC123227631	10	12	793	88,856.8	6.92	0.375	108.15	Plasma membrane
6	XP_044509128.1	*MiHAK10*	LOC123228011	10	8	794	88,971.3	8.53	0.366	109.58	Plasma membrane
7	XP_044492982.1	*MiHAK12*	LOC123216588	5	9	837	93,080	5.78	0.373	107.35	Plasma membrane
8	XP_044490120.1	*MiHAK7*	LOC123214428	4	10	848	94,175.8	5.66	0.314	105.61	Plasma membrane
9	XP_044479071.1	*MiHAK8.1*	LOC123206042	Unknown	9	775	86,717.8	7.53	0.389	110.76	Plasma membrane
10	XP_044465942.1	*MiHAK4*	LOC123196123	14	9	784	87,470.5	8.58	0.448	111.39	Plasma membrane
11	XP_044466397.1	*MiHAK8*	LOC123196431	14	8	775	86,321.2	8.15	8.15	109.25	Plasma membrane
12	XP_044512380.1	*MiHAK5*	LOC123230292	12	7	804	89,457.1	8.57	0.301	108.93	Plasma membrane
13	XP_044501404.1	*MiHAK5.3*	LOC123222612	8	8	779	87,506.1	8.04	0.221	105.56	Plasma membrane
14	XP_044502490.1	*MiHAK5.1*	LOC123223391	8	8	772	86,120.4	6.72	0.279	106	Plasma membrane
15	XP_044492757.1	*MiHAK5.2*	LOC123216406	5	8	780	87,931.8	7.54	0.233	104.79	Plasma membrane
16	XP_044467837.1	*MiHAK1*	LOC123197584	15	10	737	81,919.4	8.68	0.522	112.55	Plasma membrane
17	XP_044461170.1	*MiHAK2*	LOC123192615	12	11	793	88738.3	6.54	0.358	107.53	Plasma membrane
18	XP_044503858.1	*MiKEA2*	LOC123224297	8	21	1,208	130,774	4.92	0.093	105.38	Plasma membrane
19	XP_044464762.1	*MiKEA4*	LOC123195181	13	20	573	61,984.8	6.09	0.635	120.75	Plasma membrane
20	XP_044472009.1	*MiKEA6*	LOC123200724	17	20	579	62,521.5	5.39	0.659	124.58	Plasma membrane
21	XP_044490612.1	*MiKEA3*	LOC123214717	4	19	806	88,152.1	5.47	0.283	109.98	Plasma membrane
22	XP_044506989.1	*MiKEA5*	LOC123226529	9	20	577	62,700.8	6.33	0.628	122.43	Plasma membrane
23	XP_044502477.1	*MiTPK4*	LOC123223382	8	3	347	43,193.9	8.82	0.26	107	Plasma membrane
24	XP_044496794.1	*MiTPK6*	LOC123219111	6	2	426	47,442.5	8.96	0.094	97.65	Plasma membrane
25	XP_044464971.1	*MiTPK1.1*	LOC123195341	1	3	353	39,058.5	5.74	0.265	111.53	Plasma membrane

(Continued)

10.3389/fpls.2022.1102201

TABLE 1 Continued

Sr. no.	Accession no.	Gene names	Gene ID	Chr. No.	Exon	Protein length (aa)	Molecular weight (Da)	pI	Gravy	Aliphatic index	Subcellular localization
26	XP_044473453.1	*MiTPK1*	LOC123201916	18	5	352	39,220.9	8.31	0.129	99.94	Plasma membrane
27	XP_044503723.1	*MiTPK5*	LOC123224194	8	3	341	43,864.1	6.1	0.136	105.96	Plasma membrane
28	XP_044462393.1	*MiTPK6.1*	LOC123193455	12	3	388	46,894.7	8.94	0.154	98.11	Plasma membrane
29	XP_044496428.1	*MiTPK1.2*	LOC123218838	6	2	340	38,379.9	8.4	0.264	107.53	Plasma membrane
30	XP_044506365.1	*MiSKOR1*	LOC123226010	9	13	795	94,330.7	6.07	−0.058	−0.058	Plasma membrane
31	XP_044464305.1	*MiAKT1*	LOC123194882	13	11	892	100,795	7.33	−0.156	94.43	Plasma membrane
32	XP_044464698.1	*MiAKT6*	LOC123195143	13	12	866	97,411.9	7.86	−0.105	95.61	Plasma membrane
33	XP_044471608.1	*MiAKT1.1*	LOC123200477	17	10	873	98,247.9	8.18	−0.091	96.52	Plasma membrane
34	XP_044464837.1	*MiKAT3*	LOC123195242	13	13	622	71,509.6	7.36	−0.084	97.62	Plasma membrane
35	XP_044490239.1	*MiAKT2*	LOC123214504	4	11	838	95,849.4	6.49	−0.162	95.27	Plasma membrane
36	XP_044461595.1	*MiKAT2*	LOC123192942	12	10	823	90,520.2	6.23	−0.238	88.61	Plasma membrane
37	XP_044496295.1	*MiKAT2.1*	LOC123218753	6	11	786	89,417.4	6.44	−0.167	90.77	Plasma membrane

10.3389/fpls.2022.1102201

genes in *M. indica*. The results of this analysis demonstrated that the number of exons varies from 2 to 20. In HKTs, there were only two exons. In HAKs, the number of exons ranged from 7 to 12. Similarly, 19–21, 2–5, and 10–13 exons were predicted in KEA, TPK, and Shaker families, respectively (Figure 2A). In some regions, CDS are less in concentration and illustrated in green color in the legend. Similarly, conserved motifs were predicted and visualized by using MEME suite. A total of 10 conserved motifs were predicted in all proteins. Among these, motifs 1–8 and 10 were present in all proteins. Motif 9 was detected in only KEAs, TPKs, and Shakers. This motif possesses the characteristic GYGD motif, which acts as a selectivity filter for K^+ (Figure 2B).

3.4 Conserved domains and chromosomal distribution PTGs

The conserved domains present in the PTG proteins in *M. indica* were analyzed using NCBI-CDD. After this, further analysis was performed in phases and visualized by using the software Tbtools. All the members of PTGs possess conserved domains like (the K_trans superfamily, ions_trans_2, Na_H_Exchanger, KHA, FRQI superfamily, PLN00149, and PLN00151). Not only the most conserved domains are visualized but less conserved domains are also shown and can be visualized. The least conserved domains are from EFhs Superfamily, which revealed that some proteins function distantly (Figure 3).

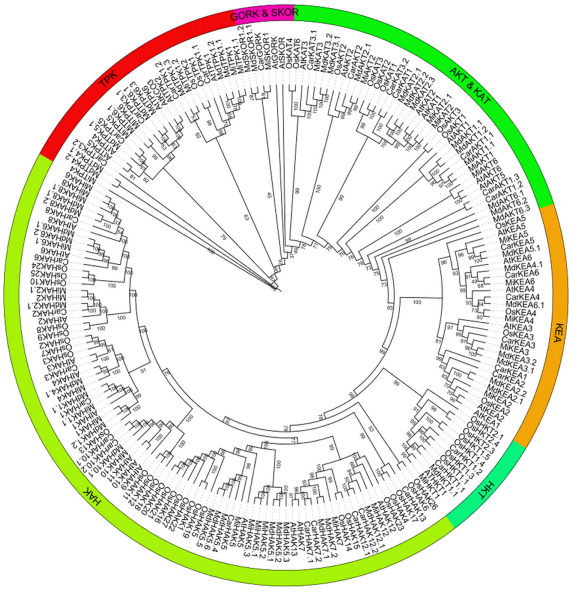


FIGURE 1
Phylogenetics analysis of K^+ transporting proteins in *Mangifera indica* (Mi), *Oryza sativa* (Os), *Arabidopsis thaliana* (At), *Cicer arietinum* (Car), and *Malus domestica* (Md). Tree was generated by MEGA-7, using 1,000 bootstrap values with the neighbor-joining method.

In *M. indica*, 36 of 37 genes were mapped on 14 (out of 18) chromosomes (Supplementary Figure 1). The locus of *MiHAK8.1* gene was not found. As of now, this gene has not been assigned to any linkage groups on scaffolds. Hence, its location is not displayed on the map. There were no PTGs on chromosomes 3, 7, 11, and 16. Chromosome 8 contains the maximum number (5) of genes, while only one gene was present on chromosomes 1, 2, 15, and 18. Almost 40% of members (13 out of 36) were present on chromosomes 8, 12, and 13 collectively (Supplementary Figure 1).

3.5 *Cis*-regulatory elements analysis

Gene expression is controlled by *cis*-regulatory sequences, such as enhancers and promoters, which play crucial roles in modulating the development and physiology of an organism. In the current study, we have identified 60 types of *cis*-regulatory elements in the 2,000-bp promoter sequences of 37 PTGs in *M. indica*. The promoter sequence of each gene was found to be rich in these regulatory elements. Most abundant of them all were found to be TATA-Box (TATA-box is capable of defining the direction of transcription and also indicates which DNA strand should be read by the transcriptional machinery), CAAT-box (CAAT-box serves as a marker for the binding site of the transcription factor for RNA), O2-site (assists in regulating two transcription factors, those associated with the metabolism of carbon and amino acids as well as the resistance to abiotic stress), AT-rich element (a replication complex is formed at this specific site and DNA synthesis is initiated), CAT-box (*cis*-acting regulatory element related to meristem expression), ARE (regulatory element essential for anaerobic induction), MRE (MYB binding site involved in light responsiveness), MBS (MYB binding site involved in drought-response), BOX-4 (conserved DNA module involved in light responsiveness), TCA-element (regulatory element involved in salicylic acid responsiveness), ABRE (Abscisic Acid-Responsive Element), and P-box (gibberellin-responsive element); in addition to these, other regulatory elements were also identified, which were not present in the promoter sequence of all the genes and were found to be present in a very small number but could have a potentially significant role in the life of *M. indica* plant, which includes Circadian, MREG-box, LTRGCN4_motif, TCT-motif, CGTCA-motif, TGACG-motif, G-Box, GA-motif, chs-CMA2aAT1-motif, ATCT-motif, LAMP-element, GATA-motif, Box III, AE-box, LS7TCCC-motif, TC-rich repeats, chs-Unit 1, m1ATC-motif, A-box, 3-AF1 binding site, TGA-element, TATC-box, I-box, GARE-motif, GT1-motif, AuxRR-core, ACECCAAT-box, SARECAG-motif, AT-rich sequence, GC-motif, RY-element, GTGGC-motif, chs-CMA1a, Box II, Gap-

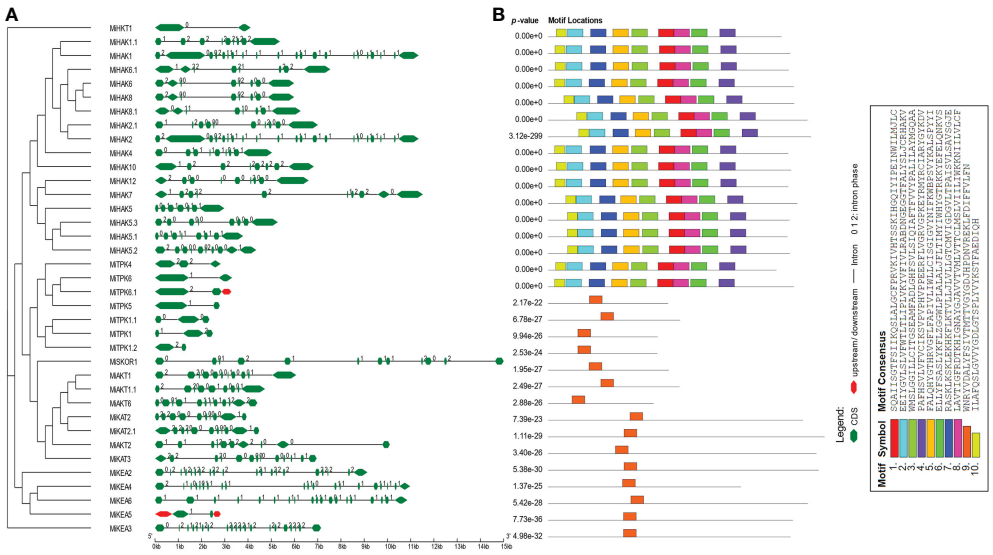


FIGURE 2
Genomic and proteomic features of PTGs. **(A)** Intron/exon architecture of PTGs in *Mangifera indica*. The green lines indicate introns. The upstream and downstream regions of genes are represented in red color. The exons are represented by green shapes. **(B)** Prediction of conserved motifs PTG proteins from *M. indica*. The conserved motifs are shown in this figure along with their motif details.

FIGURE 3
Most conserved domain of K^+ transporting proteins in *Mangifera indica*.

box, MBSIHD-Zip 1, MSA-like, Sp1, ACA-motif, TGA-box, L-box, and WUN-motif (Figure 4).

3.6 Expression profiling based on NGS data analysis of drought stress response and fruit development

To check the gene expression profile of PTGs in *M. indica* leaflets and roots under drought stress, the gene expression data of *M. indica* were downloaded from the NCBI-SRA database (Figure 5A). The differentially regulated genes were either upregulated in both leaves and roots (*MiKAT2.1*, *MiHAK2*, *MiTPK5*, *MiHKT1*, *MiHKT7*, *MiTPK1.1*, *MiHAK12*, *MiHAK1.1*, *MiAKT6*, and *MiHKT1*), upregulated in leaves and downregulated in the root (*MiAKT1*, *MiKEA2*, *MiAKT2*, *MiKEA3*, *MiKEA5*, *MiHAK8*, *MiHAK5.2*, *MiHAK5.1*, *MiKAT3*, and *MiTPK4*), and upregulated in root and downregulated in leaves (*MiHAK5*, *MiTPK6.1*, *MiAKT1.1*, *MiTPK1.2*, *MiHAK1*, *MiHAK6.1*, and *MiHAK2.1*), while the rest of the genes were non-responsive (Figure 5A). Interestingly, *MiTPK1.2*, *MiHAK1*, *MiHAK2.1*, *HAK6.1*, and *MiAKT1.1* were most upregulated in roots, and *MiKEA2*, *MiAKT2*, and *MiAKT1* were upregulated in leaves.

RNA-seq data were also used to analyze the gene expression profile of PTGs during the five stages of fruit ripening in *M. indica*. SRA data for the bioProject (PRJNA797728) were retrieved from the NCBI-SRA database. Gene expression profiles were divided into two groups, i.e., those genes that were upregulated in five different stages of fruit ripening (*MiAKT6*, *MiHAK1.1*, *MiKAT2*, *MiKAT2.1*, *MiHKT1*, *MiTPK1.1*, *MiHAK7*, and *MiHAK12*) and those genes that

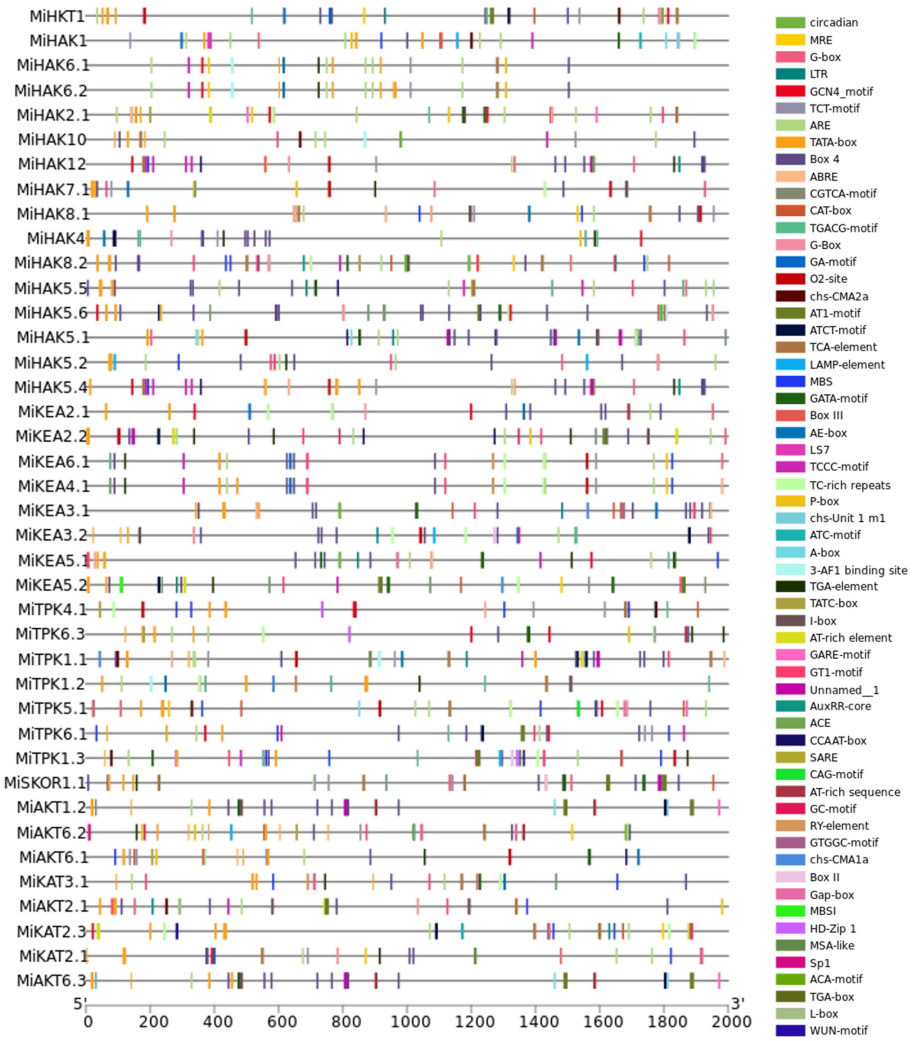


FIGURE 4
Cis-regulatory elements within 2,000-bp promoter region of PTGs.

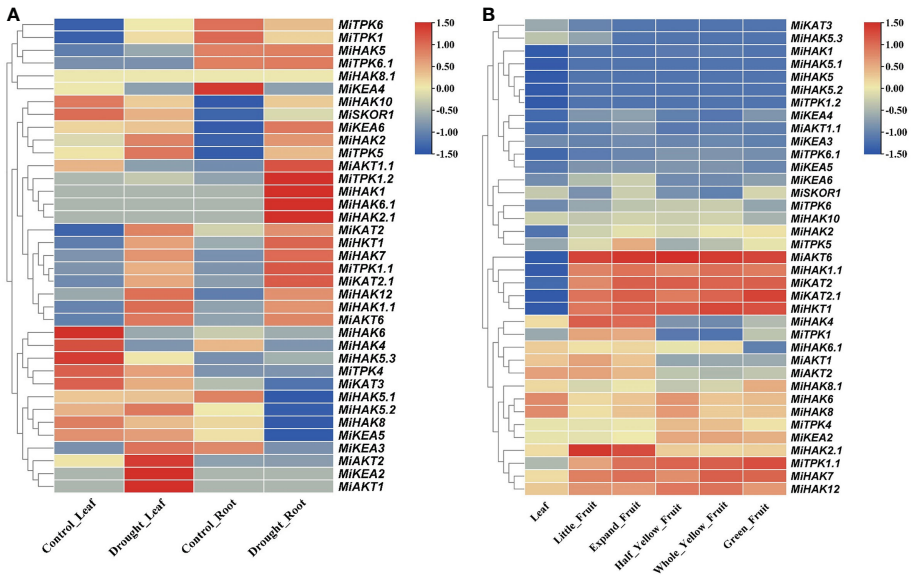


FIGURE 5
(A) Profiles of PTG expression in leaves and roots of *Mangifera indica*. The expression values [log2(FPKM)] are shown by the colored bar. Red denotes high expression, yellow denotes no expression, and blue represents low expression. **(B)** Profiles of PTG expression during fruit development. The expression values [log2(FPKM)] are shown by the colored bar. Red denotes high expression, whereas blue denotes low expression.

were highly downregulated in leaves and fruit ripening according to FPKM values (*MiHAK1*, *MiHAK5.1*, *MiHAK5*, *MiHAK5.2*, and *MiTPK1.2*), and the rest of the genes showed non-responsive or non-significant variations of expression (Figure 5B).

3.7 Relative expression profiling of PTGs using real-time RT-qPCR

Real-time amplification (qRT-PCR) was performed to confirm the expression of nine selected PTGs in *M. indica* leaves. These genes were selected based on their expression values in drought stress response (Figure 5A) and higher expression during fruit development (Figure 5B). In response to drought stress, nine genes, including *MiAKT6*, *MiHAK1.1*, *MiKAT2.1*, *MiHKT1*, *MiTPK1.1*, *MiHAK1*, *MiHAK5.1*, *MiHAK5*, and *MiHAK5.2*, were selected for qPCR-based quantification for 7 days, and samples were taken at the second, fourth, and seventh days. Under drought stress, *MiHAK1* was regulated differently. Although according to RNA-seq analysis *MiHAK1* was downregulated in response to drought stress, it was found to be upregulated up to 3.5-fold in qPCR. Similarly, *MiHAK1.1*, *MiHAK5.2*, *MiTPK1.1*, *MiKAT2.1*, and *MiAKT6* were significantly upregulated after drought stress. On the contrary, *MiHAK5*, *MiHAK5.1*, and *MiHKT1* were significantly downregulated (Figure 6). There was a significant correlation between the results of qPCR and RNA-seq data. However, some variations were also observed. Though similar plant growth and stress conditions were applied, such variations are potentially associated with sampling time in a day, plant age, or even genotypic differences.

4 Discussion

4.1 Potassium transport system is highly conserved in *M indica* and other plants

The current study identified 37 potential genes of the K^+ transport system in *M. indica* genome. The number of PTGs is more or less similar in major plant groups represented by *V. vinifera*, *Glycine max*, *Triticum aestivum*, *Vigna radiata*, *C. cajan*, and *A. thaliana* (Cuéllar et al., 2010; Isayenkov et al., 2011; Chen et al., 2017; Azeem et al., 2018; Azeem et al., 2021a; Azeem et al., 2021b; Ge et al., 2020; Siddique et al., 2021; Cai et al., 2021). It not only indicates the evolutionary conservation of PTGs in these plants but also predicts the existence of similar genetic mechanisms for K^+ in other fruit trees. A total of 202 PTG protein sequences from five species—*A. thaliana*, *M. indica*, *M. domestica*, *O. sativa*, and *C. arietinum*—were divided into six groups according to phylogenetic analyses (HAK, HKT, KEA, AKT and KAT, GORK and SKOR, and TPK). In contrast, the phylogenetic relationship analysis showed that *PTGs* were more closely related to *MdPTGs* than *AtPTGs*. This outcome confirmed the finding that apple and mango also showed a closer association than *Arabidopsis*. It is also supported

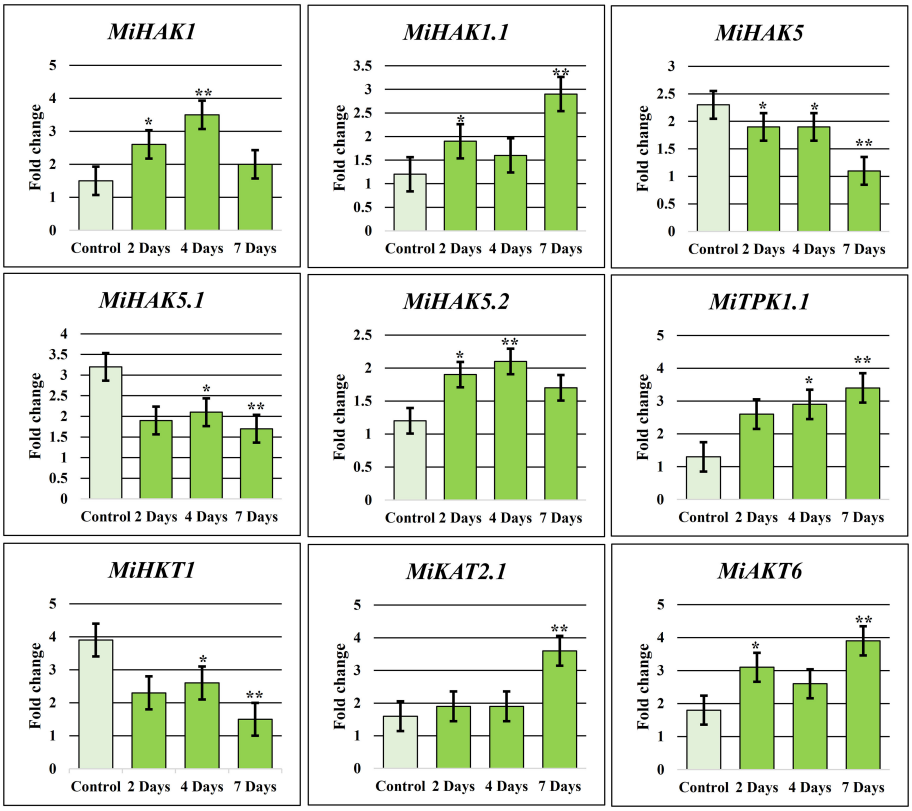


FIGURE 6
K^+ transporting gene family relative qRT-PCR in response to drought. To produce an objective average value, the experiment was triplicated. In untreated plants, each gene had a default expression value of 1. Bars have been placed on each column to illustrate the standard error. *denotes significant differences between environmental stressors and the control (p 0.05), whereas **denotes extremely significant differences (p 0.001).

by the fact that various sequence features were common in *M. indica* and other species. For example, the transmembrane domains of MiHKT1.1 and AtHKT1;1 are eight. Similarly, the presence of the consensus motif GVVYGDLGTSPLY is a characteristic feature of HAK transporters (Cheng et al., 2018). This motif was present with a minor modification as GVVYGDLG(I/T)SPLY. This variation of a motif is also present in HAK proteins in *A. thaliana* and *Camellia sinensis* (Yang et al., 2020). KEA members of both *M. indica* and *A. thaliana* species showed conservation of "FLLFxxGLE and GEFAFVxxxxA" motifs. Previously, it has been established that a divergence in the amino acid sequences of proteins is related to the functional divergence of proteins and vice versa (Sangar et al., 2007). Among K^+ channels, the presence of ANK, Ion_trans_2, and KHA domains in the Shaker proteins, similar to *Arabidopsis* K^+ channels, suggests their functional similarities (Keisham et al., 2018). Likewise, the occurrence of the significantly conserved residues, i.e., GYGD and RSXpS/pTXP, in MiTPKs indicates potential functional similarities in *M. indica* and other plants (Ge et al., 2020; Siddique et al., 2021).

4.2 Potassium transport system is involved in drought stress response in *M. indica*

As one of the most abundant cations in plant cells, K^+ performs a dominant role in various biological processes of plants (Ragel et al., 2019). The influx/efflux of K^+ (mediated by channel and transporter proteins) regulates its concentration in the plant body. K^+ channels and transporters are important contributors to plant growth and development (Sharma et al., 2013). In the current study, members of HAK, HKT, TPK, and Shaker families were differentially regulated by drought stress. These results were more or less in agreement with RNA-seq and qPCR analyses. Previously, the role of potassium channels and transporters has been documented in the drought stress response of various plants (Ahmad et al., 2016a; Ahmad et al., 2016b; Chen et al., 2017; Hassan et al., 2017; Cai et al., 2019; Qi et al., 2019; Singh et al., 2021). The overexpression of the *OsAKT1* and *OsHAK1* significantly affects potassium nutrition and

drought stress tolerance of rice (Ahmad et al., 2016a; Chen et al., 2017). Our results also proposed similar roles of K^+ transporters. Overexpression of the potassium channel TPKb in small vacuoles confers osmotic and drought tolerance to rice (Ahmad et al., 2016b). Taken together, these findings suggest a potential role of PTGs in the stress response of the *M. indica* tree.

4.3 *M. indica* fruit development involves differential regulation of PTGs

The essential nutrients (N, P, and K) promote an increase in biomass yield by improving root growth, activating cellular enzymes, ameliorating photosynthesis, conserving energy, and positively affecting other fundamental processes in the plant body. Potassium sources can influence the K^+ release and fruit yield of *M. indica* (Wang et al., 2022). It emphasizes the importance of potassium uptake and transport system from root to leaves and fruit. The current study reports significant differential regulation of K^+ transport during five stages of mango fruit development. Previously, it is reported that supplemental foliar applications of K^+ can improve fruit quality in *Cucumis melo* (Lester et al., 2005), *Solanum lycopersicum* (Liu et al., 2021), *C. sinensis* (Wu et al., 2021), and *M. indica* (Sarker and Rahim, 2013). The findings of the current study complement the potential role of PTGs in improving mango fruit growth, development, and quality.

Conclusion

In the current study, the K^+ transport system was characterized in *M. indica*. There are 37 potential members of PTGs in this plant. Among these, 22 genes can be classified as K^+ transporters, and the remaining 15 genes are K^+ channels. Moreover, the analysis performed including conserved domain and TM domain prediction, motif analysis, phylogenetic analysis, and genetic structure display, and *cis*-regulatory elements predicted in the promoter regions revealed the close relation of PTGs in several plant species. Differentially expressed genes *MiHKT7*, *MiHAK5.1*, *MiHAK5*, *MiTPK1.1*, and *MiHAK12* were responsive to drought stress. Moreover, *MiHKT7*, *MiHAK5.1*, *MiHAK5*, *MiTPK1.1*, and *MiHAK12* genes were also differentially expressed during different growth stages of mango fruit development. To gain depth of information about the K^+ transport system, these genes can be selected for further genomic and biotechnological study and to improve stress response and fruit quality in mango.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

Author contributions

Conceptualization: FA. Data curation: TL and MW. Formal analysis: TL and AR. Funding acquisition: FA and TL. Investigation: TL and MR. Methodology, MW and AR. Project administration: FA. Resources: FA. Software: MW and AR. Supervision: FA. Validation: FA, TL, and HM. Visualization: MW. Writing—original draft: TL. Writing—review and editing: FA and HM. All authors contributed to the article and approved the submitted version.

Funding

This research was funded by the Hainan Provincial Natural Science Foundation of China (grant number 321RC617).

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.1102201/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Distribution of PTGs on *M. indica* on chromosomes. The 37 PTGs were mapped onto all chromosomes except 3, 7, 11, and 16. Each gene is represented by a specific color.

References

Ahmad, I., Devonshire, J., Mohamed, R., Schultze, M., and Maathuis, F. J. M. (2016a). Overexpression of the potassium channel TPKb in small vacuoles confers osmotic and drought tolerance to rice. *New Phytol.* 209, 1040–1048. doi: 10.1111/nph.13708

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

Aranda-Sicilia, M. N., Cagnac, O., Chanroj, S., Sze, H., Rodríguez-Rosales, M. P., and Venema, K. (2012). Arabidopsis KEA2, a homolog of bacterial KefC, encodes a K+/H+ antiporter with a chloroplast transit peptide. *Biochim. Biophys. Acta -Biomembranes.* 1818 (9), 2362–2371. doi: 10.1016/j.bbamem.2012.04.011

Assaha, D. V., Mekawy, A. M. M., Ueda, A., and Saneoka, H. (2015). Salinity-induced expression of HKT may be crucial for na+ exclusion in the leaf blade of huckleberry (Solanum scabrum mill.), but not of eggplant (Solanum melongena l.). *Biochem. Biophys. Res. Commun.* 460 (2), 416–421. doi: 10.1016/j.bbrc.2015.03.048

Azeem, F., Ahmad, B., Atif, R. M., Ali, M. A., Nadeem, H., Hussain, S., et al. (2018). Genome-wide analysis of potassium transport-related genes in chickpea (Cicer arietinum l.) and their role in abiotic stress responses. *Plant Mol. Biol. Rep.* 36, 451–468. doi: 10.1007/s11105-018-1090-2

Azeem, F., Hussain, M., Hussain, S., Zubair, M., Nadeem, H., Ali, M. A., et al. (2021a). Genome-wide analysis and expression profiling of potassium transport related genes in solanum tuberosum. *Pakistan J. Agric. Sci.* 58, 81–94. doi: 10.3390/plants11010002

Azeem, F., Ijaz, U., Ali, M. A., Hussain, S., Zubair, M., Manzoor, H., et al. (2021b). Genome-wide identification and expression profiling of potassium transport-related genes in vigna radiata under abiotic stresses. *Plants* 11 (1), 2. doi: 10.3390/plants11010002

Azeem, F., Zameer, R., Rashid, M. A. R., Rasul, I., Ul-Allah, S., Siddique, M. H., et al. (2022). Genome-wide analysis of potassium transport genes in gossypium raimondii suggest a role of GrHAK/KUP/KT8, GrAKT2. 1 and GrAKT1. 1 in response to abiotic stress. *Plant Physiol. Biochem.* 170, 110–122. doi: 10.1016/j.plaphy.2021.11.038

Bahadur, I., Maurya, R., Roy, P., and Kumar, A. (2019). "Potassium-solubilizing bacteria (KSB): a microbial tool for K-solubility, cycling, and availability to plants," in *Plant growth promoting rhizobacteria for agricultural sustainability* (Springer), 257–265. doi: 10.1007/978-981-13-7553-8_13

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., et al. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 37 (suppl_2), W202–W208. doi: 10.1093/nar/gkp335

Cai, K., Gao, H., Wu, X., Zhang, S., Han, Z., Chen, X., et al. (2019). The ability to regulate transmembrane potassium transport in root is critical for drought tolerance in barley. *Int. J. Mol. Sci.* 20, 4111. doi: 10.3390/ijms20174111

Cai, K., Zeng, F., Wang, J., and Zhang, G. (2021). Identification and characterization of HAK/KUP/KT potassium transporter gene family in barley and their expression under abiotic stress. *BMC Genomics* 22 (1), 1–14. doi: 10.1186/s12864-021-07633-y

Celestin, A. (2019). Bridging the gap between smallholder farmers and market access through agricultural value chain development in Haiti. *Agronomy*.

Cheng, X., Liu, X., Mao, W., Zhang, X., Chen, S., Zhan, K., et al. (2018). Genome-wide identification and analysis of HAK/KUP/KT potassium transporters gene family in wheat (Triticum aestivum l.). *Int. J. Mol. Sci.* 19, 3969. doi: 10.3390/ijms19123969

Chen, G., Liu, C., Gao, Z., Zhang, Y., Jiang, H., Zhu, L., et al. (2017). OsHAK1, a high-affinity potassium transporter, positively regulates responses to drought stress in rice. *Front. Plant Sci.* 8, 1885. doi: 10.3389/fpls.2017.01885

Cleveland, R. P. (2012). Second report on the state of plant genetic resources for food and agriculture in guyana. *Agronomy Science*. (EC Demerara: National Agricultural Research and Extension Institute Mon Repos)

Council, N. R. (1993). *Sustainable agriculture and the environment in the humid tropics* (Washington, DC: The National Academies Press). doi: 10.17226/1985.

Cuéllar, T., Pascaud, F., Verdeil, J.-L., Torregrosa, L., Adam-Blondon, A.-F., Thibaud, J.-B., et al. (2010). A grapevine shaker inward k(+) channel activated by the calcineurin b-like calcium sensor 1-protein kinase CIPK23 network is expressed in grape berries under drought stress conditions. *Plant J.* 61, 58–69. doi: 10.1111/j.1365-313X.2009.04029.x

Etesami, H., Emami, S., and Alikhani, H. A. (2017). Potassium solubilizing bacteria (KSB):: Mechanisms, promotion of plant growth, and future prospects a review. *J. Soil Sci. Plant Nutr.* 17 (4), 897–911. doi: 10.4067/S0718-95162017000400005

FAOSTAT (2022). *Production of mangoes, mangosteens, and guavas in 2020, Crops/Regions/World list/Production quantity (pick lists)* (UN Food and Agriculture Organization, Corporate Statistical Database (FAOSTAT). (Rome, Italy: Viale delle Terme di Caracalla)

Farooq, M., Hussain, M., and Siddique, K. H. (2014). Drought stress in wheat during flowering and grain-filling periods. *Crit. Rev. Plant Sci.* 33 (4), 331–349. doi: 10.1080/07352689.2014.875291

Galán Saúco, V. (2002). "Mango production and world market: Current situation and future prospects," in VII international mango symposium. *ISHS Acta Horticulturae.* 645, 107–116.

Garg, V. K., Avashthi, H., Tiwari, A., Jain, P. A., Ramkete, P. W., Kayastha, A. M., et al. (2016). MFPPI–multi FASTA ProtParam interface. *Bioinformation* 12 (2), 74. doi: 10.6026/97320630012074

Ge, M., Zhang, L., Ai, J., Ji, R., He, L., Liu, C., et al (2020). Effect of heat shock and potassium sorbate treatments on gray mold and postharvest quality of 'XuXiang' kiwifruit. *Food Chem.* 324, 126891. doi: 10.1016/j.foodchem.2020.126891.

González, W., Valdebenito, B., Caballero, J., Riadi, G., Riedelsberger, J., Martínez, G., et al. (2015). K2P channels in plants and animals. *Pflügers. Archiv-European. J. Physiol.* 467 (5), 1091–1104. doi: 10.1007/s00424-014-1638-4

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 (D1), D1178–D1186. doi: 10.1093/nar/gkr944

Guo, A.-Y., Zhu, Q.-H., Chen, X., and Luo, J.-C. (2007). GSDS: a gene structure display server. *Yi. chuan=. Hereditas.* 29 (8), 1023–1026. doi: 10.1360/yc-007-1023

Hasanuzzaman, M., Bhuyan, M. B., Nahar, K., Hossain, M. S., Mahmud, J. A., Hossen, M. S., et al. (2018). Potassium: a vital regulator of plant responses and tolerance to abiotic stresses. *Agronomy*. 8, 31. doi: 10.3390/agronomy8030031

Hossain, A., Skalicky, M., Brestic, M., Maitra, S., Ashraful Alam, M., Syed, M. A., et al. (2021). Consequences and mitigation strategies of abiotic stresses in wheat (Triticum aestivum l.) under the changing climate. *Agronomy* 11 (2), 241. doi: 10.3390/agronomy11020241

Hassan, M. U., Aamer, M., Chattha, M. U., Aman Ullah, M., Sulaman, S., Nawaz, M., et al. (2017). The Role of Potassium in Plants under Drought Stress: Mini Review. *J. Basic Appl. Sci.* 13, 268–271. doi: 10.6000/1927-5129.2017.13.44.

Isayenkov, S., Isner, J.-C., and Maathuis, F. J. M. (2011). Rice two-pore k+ channels are expressed in different types of vacuoles. *Plant Cell* 23, 756–768. doi: 10.1105/tpc.110.081463

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., and Madden, T. L. (2008). NCBI BLAST: a better web interface. *Nucleic Acids Res.* 36 (suppl_2), W5–W9. doi: 10.1093/nar/gkn201

Keisham, M., Mukherjee, S., and Bhatla, S. C. (2018). Mechanisms of sodium transport in plants–progresses and challenges. *Int. J. Mol. Sci.* 19, 647. doi: 10.3390/ijms19030647

Laurio, M. V. O. (2021). *Integrating vibratory membrane-based water recovery systems for sustainable food and beverage production* (Glassboro, New Jersey: Rowan University).

Lescot, M., Déhais, P., Thijs, G., Marchal, K., Moreau, Y., Van de Peer, Y., et al. (2002). PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res.* 30 (1), 325–327. doi: 10.1093/nar/30.1.325

Lester, G. E., Jifon, J. L., and Rogers, G. (2005). Supplemental foliar potassium applications during muskmelon fruit development can improve fruit quality, ascorbic acid, and beta-carotene contents. *J. Am. Soc Hortic. Sci.* 130, 649–653. doi: 10.21273/JASHS.130.4.649

Liu, J., Hu, T., Feng, P., Yao, D., Gao, F., and Hong, X. (2021). Effect of potassium fertilization during fruit development on tomato quality, potassium uptake, water and potassium use efficiency under deficit irrigation regime. *Agric. Water Manage.* 250, 106831. doi: 10.1016/j.agwat.2021.106831

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

Mehanny, M., Kroniger, T., Koch, M., Hoppstädter, J., Becher, D., Kiemer, A. K., et al. (2022). Yields and immunomodulatory effects of pneumococcal membrane vesicles differ with the bacterial growth phase. *Adv. Healthcare. Mater.* 11 (5), 2101151. doi: 10.1002/adhm.202101151

Nahar, K., Hasanuzzaman, M., Ahamed, K. U., Hakeem, K. R., Ozturk, M., and Fujita, M. (2015). "Plant responses and tolerance to high temperature stress: role of exogenous phytoprotectants," in *Crop production and global environmental issues* (Switzerland: Springer International Publishing), 385–435.

Ngamchuachit, P., Sivertsen, H. K., Mitcham, E. J., and Barrett, D. M. (2015). Influence of cultivar and ripeness stage at the time of fresh-cut processing on instrumental and sensory qualities of fresh-cut mangos. *Postharvest. Biol. Technol.* 106, 11–20. doi: 10.1016/j.postharvbio.2015.03.013

Qi, J., Sun, S., Yang, L., Li, M., Ma, F., and Zou, Y. (2019). Potassium uptake and transport in apple roots under drought stress. *Hortic. Plant J.* 5, 10–16. doi: 10.1016/j.hpj.2018.10.001

Ragel, P., Raddatz, N., Leidi, E. O., Quintero, F. J., and Pardo, J. M. (2019). Regulation of k+ nutrition in plants. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.00281

Riedelsberger, J., Miller, J. K., Valdebenito-Maturana, B., Piñeros, M. A., González, W., and Dreyer, I. (2021). Plant HKT channels: an updated view on structure, function and gene regulation. *Int. J. Mol. Sci.* 22 (4), 1892. doi: 10.3390/ijms22041892

Sahi, N., Mostajeran, A., and Ghanadian, M. (2022). Changing in the production of anticancer drugs (vinblastine and vincristine) in catharanthus roseus (L.) g. don by potassium and ascorbic acid treatments. *Plant. Soil Environ.* 68 (1), 18–28. doi: 10.17221/121/2021-PSE

Sangar, V., Blankenberg, D. J., Altman, N., and Lesk, A. M. (2007). Quantitative sequence-function relationships in proteins based on gene ontology. *BMC Bioinf.* 8, 294. doi: 10.1186/1471-2105-8-294

Sarker, B. C., and Rahim, M. (2013). Yield and quality of mango (Mangifera indica l.) as influenced by foliar application of potassium nitrate and urea. *Bangladesh. J. Agric. Res.* 38, 145–154. doi: 10.3329/bjar.v38i1.15201

Schultz, J., Copley, R. R., Doerks, T., Ponting, C. P., and Bork, P. (2000). SMART: a web-based tool for the study of genetically mobile domains. *Nucleic Acids Res.* 28 (1), 231–234. doi: 10.1093/nar/28.1.231

Sharma, T., Dreyer, I., and Riedelsberger, J. (2013). The role of k+ channels in uptake and redistribution of potassium in the model plant arabidopsis thaliana. *Front. Plant Sci.* 4. doi: 10.3389/fpls.2013.00224

Siddique, M. H., Babar, N. I., Zameer, R., Muzammil, S., Nahid, N., Ijaz, U., et al. (2021). Genome-wide identification, genomic organization, and characterization of potassium transport-related genes in cajanus cajan and their role in abiotic stress. *Plants* 10, 2238. doi: 10.3390/plants10112238

Singh, S., Kumar, V., Parihar, P., Dhanjal, D. S., Singh, R., Ramamurthy, P. C., et al. (2021). Differential regulation of drought stress by biological membrane transporters and channels. *Plant Cell Rep.* 40, 1565–1583. doi: 10.1007/s00299-021-02730-4

Singh, K., and Kushwaha, C. (2006). Diversity of flowering and fruiting phenology of trees in a tropical deciduous forest in India. *Ann. Bot.* 97 (2), 265–276. doi: 10.1093/aob/mcj028

Snowden, M. C., Ritchie, G. L., Simao, F. R., and Bordovsky, J. P. (2014). Timing of episodic drought can be critical in cotton. *Agron. J.* 106 (2), 452–458. doi: 10.2134/agronj2013.0325

Sun, J., Cao, H., Cheng, J., He, X., Sohail, H., Niu, M., et al. (2018). Pumpkin CmHKT1; 1 controls shoot na+ accumulation *via* limiting na+ transport from rootstock to scion in grafted cucumber. *Int. J. Mol. Sci.* 19 (9), 2648. doi: 10.3390/ijms19092648

Sze, H., and Chanroj, S. (2018). Plant endomembrane dynamics: studies of K+/H + antiporters provide insights on the effects of pH and ion homeostasis. *Plant Physiol.* 177 (3), 875–895. doi: 10.1104/pp.18.00142

Ul Hassan, M., Rasool, T., Iqbal, C., Arshad, A., Abrar, M., Abrar, M. M., et al. (2021). Linking plants functioning to adaptive responses under heat stress conditions: a mechanistic review. *J. Plant Growth Regul.* 41, 2596"C2613. doi: 10.1007/s00344-021-10493-1

Voelker, C., Gomez-Porras, J., Becker, D., Hamamoto, S., Uozumi, N., Gambale, F., et al. (2010). Roles of tandem-pore k+ channels in plants–a puzzle still to be solved. *Plant Biol.* 12, 56–63. doi: 10.1111/j.1438-8677.2010.00353.x

Wang, J., Ding, Z., AL-Huqail, A. A., Hui, Y., He, Y., Reichman, S. M., et al. (2022). Potassium source and biofertilizer influence K release and fruit yield of mango (Mangifera indica l.): A three-year field study in sandy soils. *Sustainability* 14 (15), 9766. doi: 10.3390/su14159766

Wang, C., Yamamoto, H., Narumiya, F., Munekage, Y. N., Finazzi, G., Szabo, I., et al. (2017). Fine-tuned regulation of the K+/H+ antiporter KEA 3 is required to optimize photosynthesis during induction. *Plant J.* 89 (3), 540–553. doi: 10.1111/tpj.13405

Waters, S., Gilliham, M., and Hrmova, M. (2013). Plant high-affinity potassium (HKT) transporters involved in salinity tolerance: structural insights to probe differences in ion selectivity. *Int. J. Mol. Sci.* 14 (4), 7660–7680. doi: 10.3390/ijms14047660

Wu, S., Zhang, C., Li, M., Tan, Q., Sun, X., Pan, Z., et al. (2021). Effects of potassium on fruit soluble sugar and citrate accumulations in cara cara navel orange (Citrus sinensis l. osbeck). *Sci. Hortic. (Amsterdam).* 283, 110057. doi: 10.1016/j.scienta.2021.110057

Yang, T., Lu, X., Wang, Y., Xie, Y., Ma, J., Cheng, X., et al. (2020). HAK/KUP/KT family potassium transporter genes are involved in potassium deficiency and stress responses in tea plants (Camellia sinensis l.): expression and functional analysis. *BMC Genomics* 21, 556. doi: 10.1186/s12864-020-06948-6

Yao, R., Huang, X., Cong, H., Qiao, F., Cheng, Y., and Chen, Y. (2022). Selection and identification of a reference gene for normalizing real-time PCR in mangos under various stimuli in different tissues. *Horticulturae* 8, 882. doi: 10.3390/horticulturae8100882

Ye, C.-Y., Yang, X., Xia, X., and Yin, W. (2013). Comparative analysis of cation/proton antiporter superfamily in plants. *Gene* 521 (2), 245–251. doi: 10.1016/j.gene.2013.03.104

Zhu, X., Pan, T., Zhang, X., Fan, L., Quintero, F. J., Zhao, H., et al. (2018). K+ efflux antiporters 4, 5, and 6 mediate pH and k+ homeostasis in endomembrane compartments. *Plant Physiol.* 178 (4), 1657–1678. doi: 10.1104/pp.18.01053

Zikki, K. A. F. (2020). *Quality of bell pepper (Capsicum annuum l.) affected by drought condition: a thesis presented in the partial fulfilment of the requirements for the degree of master of philosophy at Massey university, palmerston north, new Zealand* (New Zealand: Massey University).

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 30 January 2023
DOI 10.3389/fpls.2023.1101665

OPEN ACCESS

EDITED BY
Galal Bakr Anis,
Field Crops Research Institute, Egypt

REVIEWED BY
Sandhya Verma,
Shri Vaishnav Vidyapeeth Vishwavidyalaya, Indore, India
Longbiao Guo,
China National Rice Research Institute (CAAS), China

*CORRESPONDENCE
Muqing Zhang
zmuqing@163.com
Wei Yao
yaoweimail@163.com
Jingsheng Xu
xujingsheng@126.com

SPECIALTY SECTION
This article was submitted to Plant Biotechnology, a section of the journal Frontiers in Plant Science

RECEIVED 18 November 2022
ACCEPTED 09 January 2023
PUBLISHED 30 January 2023

CITATION
Shang H, Fang L, Qin L, Jiang H, Duan Z, Zhang H, Yang Z, Cheng G, Bao Y, Xu J, Yao W and Zhang M (2023) Genome-wide identification of the class III peroxidase gene family of sugarcane and its expression profiles under stresses. *Front. Plant Sci.* 14:1101665. doi: 10.3389/fpls.2023.1101665

Genome-wide identification of the class III peroxidase gene family of sugarcane and its expression profiles under stresses

Heyang Shang[1,2], Linqi Fang[2], Lifang Qin[2], Hongtao Jiang[1,2], Zhenzhen Duan[2], Hai Zhang[1], Zongtao Yang[1], Guangyuan Cheng[1], Yixue Bao[2], Jingsheng Xu[1,2]*, Wei Yao[2]* and Muqing Zhang[1,2]*

[1]National Engineering Research Center for Sugarcane & Guangxi Key Laboratory of Sugarcane Biology, Fujian Agriculture and Forestry University, Fuzhou, China, [2]State Key Laboratory for Conservation and Utilization of Subtropical Agro-Bioresources & Guangxi Key Laboratory of Sugarcane Biology, Guangxi University, Nanning, China

Introduction: Plant-specific Class III peroxidases (PRXs) play a crucial role in lignification, cell elongation, seed germination, and biotic and abiotic stresses.

Methods: The class III peroxidase gene family in sugarcane were identified by bioinformatics methods and realtime fluorescence quantitative PCR.

Results: Eighty-two PRX proteins were characterized with a conserved PRX domain as members of the class III PRX gene family in R570 STP. The ShPRX family genes were divided into six groups by the phylogenetic analysis of sugarcane, Saccharum spontaneum, sorghum, rice, and *Arabidopsis thaliana*. The analysis of promoter *cis*-acting elements revealed that most *ShPRX* family genes contained *cis*-acting regulatory elements involved in ABA, MeJA, light responsiveness, anaerobic induction, and drought inducibility. An evolutionary analysis indicated that ShPRXs was formed after *Poaceae* and *Bromeliaceae* diverged, and tandem duplication events played a critical role in the expansion of *ShPRX* genes of sugarcane. Purifying selection maintained the function of *ShPRX* proteins. *SsPRX* genes were differentially expressed in stems and leaves at different growth stages in *S. spontaneum*. However, *ShPRX* genes were differentially expressed in the SCMV-inoculated sugarcane plants. A qRT-PCR analysis showed that SCMV, Cd, and salt could specifically induce the expression of PRX genes of sugarcane.

Discussion: These results help elucidate the structure, evolution, and functions of the class III *PRX* gene family in sugarcane and provide ideas for the phytoremediation of Cd-contaminated soil and breeding new sugarcane varieties resistant to sugarcane mosaic disease, salt, and Cd stresses.

KEYWORDS

sugarcane, class III peroxidase, sugarcane mosaic virus, cadmium, salt stress

1 Introduction

Sugarcane, one of the critical sugar and energy crops, is often subjected to various biotic stresses, including *Sugarcane mosaic virus* (SCMV), *Sorghum mosaic virus* (SrMV), *Sugarcane streak mosaic virus* (SCSMV), and abiotic stress, including salt, heavy metal, and drought stress. Sugarcane mosaic disease, caused by SCMV, SrMV, and SCSMV, is currently one of the most severe sugarcane diseases worldwide that adversely affect the healthy and sustainable development of the sugarcane industry (Yao et al., 2017; Moradi et al., 2018; Rice et al., 2019; Rice and Hoy, 2020). Under salt stress, the growth of sugarcane is hindered, seriously affecting the quality of sugarcane and even causing a large area of yield reduction or crop failure. Soil heavy metal pollution, one of the leading environmental stresses affecting plant growth and development, is becoming the prime concern of various terrestrial ecosystems worldwide. Among heavy metals, cadmium (Cd), one of the most dangerous toxic elements for plants, inhibits various physiological processes of plants, including seed germination, seedling growth, photosynthesis, and antioxidant system (Zhu et al., 2021). The contents of chlorophyll and soluble protein decreased significantly, whereas the content of carotenoids increased significantly in the Cd-treated sugarcane. The activity of ascorbate peroxidase, peroxidase (PRX), and catalase (CAT) increased significantly (Yousefi et al., 2018). Sugarcane, one of the cultivated crops with the highest biomass and solid tolerance to Cd, can be used as a candidate crop for the phytoremediation of Cd pollution in soil (Sereno et al., 2007; Yousefi et al., 2018). Therefore, it is one of the main challenges to improve the yield and quality of sugarcane and the ability to repair Cd pollution in soil by improving the resistance of sugarcane to biotic and abiotic stresses.

The PRXs, the critical enzymes of peroxisomes, widely exist in animals, plants, and microorganisms. According to the protein structural and functional characteristics, PRXs are divided into heme PRXs and non-heme PRXs. Heme PRXs are further subdivided into animal PRXs and non-animal PRXs. According to the sequence and catalytic characteristics of proteins, non-animal heme PRXs comprise Class I, II, and III PRXs, all containing a heme group consisting of protoporphyrin IX and iron (III). Class I PRXs widely exist in most organisms, such as plants, fungi, bacteria, and protozoa. However, Class II PRXs exist in fungi, and class III PRXs only exist in plants (Piontek et al., 2001; Zamocky, 2004; Zamocky et al., 2010; Shigeto and Tsutsumi, 2016). Class III PRXs are classical plant secretory PRXs that play a crucial role in lignification, cell elongation, and seed germination (Lee, 1977; Piontek et al., 2001; Shigeto and Tsutsumi, 2016; Kidwai et al., 2020). Numerous class III PRXs were present in the cell walls, and the balance of cell wall loosening and stiffening could be precisely controlled by the antagonistic activities of class III PRXs during plant growth (Francoz et al., 2015). Class III PRXs was identified as an essential enzyme for lignin biosynthesis in plants. Coniferyl alcohol and sinapyl alcohol, precursors for the synthesis of lignin monomers, could be catalyzed by the class III PRX gene *PbPRX2* in Chinese pear fruit (Vanholme et al., 2019; Zhu et al., 2021). In addition, class III PRXs were involved in the internal browning of pineapple (Hou et al., 2022) and closely related to pollen fertility in *Gossypium hirsutum* (Chen et al., 2022).

The class III PRX genes play a vital role in responses to various biotic and abiotic stresses throughout the plant life cycle. The wheat class III PRX gene *(TaPRX-2A)* increased the activities of superoxide dismutase (SOD), PRX, and CAT to scavenge reactive oxygen species (ROS) (Su et al., 2020). In sweet potato, the B-box family transcription factor *IbBBX24* activated the expression of the class III PRX gene *IbPRX17* by binding to the promoter of *IbPRX17*, and the overexpression of *IbPRX17* significantly improved the tolerance to salt and drought stresses by scavenging ROS (Zhang et al., 2022). Overexpressed the class III PRX gene *(OsPRX38)* in *Arabidopsis thaliana* exposed to arsenic stress increased SOD, PRX, and GST activities to reduce the content of hydrogen peroxide (H_2O_2), electrolyte leakage, and malondialdehyde (Kidwai et al., 2019). In rice, the overexpression of the class III PRX *OsPRX30*, maintaining a high level of PRX activity and reducing the content of H_2O_2, reduced bacterial blight resistance (Liu et al., 2021). In *Citrus sinensis*, the class III *CsPRX* family genes, induced by salicylic acid and methyl jasmonate, were involved in citrus bacterial canker disease (Li et al., 2020).

The publication of the genome of *Saccharum* hybrid cultivar R570 (R570) BAC clone (BAC) and single tiling path (STP) has necessitated the study of the function of the sugarcane class III PRX gene. At present, the class III PRX gene family has been studied in various plant species, including *A. thaliana* (Tognolli et al., 2002), rice (Passardi et al., 2004), maize (Wang et al., 2015), potato (Yang et al., 2020), soybean (Aleem et al., 2022), *Brachypodium distachyon* (Zhu et al., 2019), and allotetraploid cotton (Duan et al., 2019). However, it has yet to be reported on the identification and characterization of the class III PRX gene family in sugarcane.

A genome-wide search was carried out on class III PRXs in R570 STP, R570 BAC, *Saccharum spontaneum* AP85-441 (*S. spontaneum* AP85-441), sorghum, *A. thaliana*, and rice. The functions of the class III PRX genes in sugarcane were analyzed concerning the gene structure, conserved motif, *cis*-acting elements, codon usage bias, and evolutionary analysis. The expression level of *ShPRXs* was studied under SCMV, Cd, and salt stress response. The results from this study help elucidate the structure, evolution, and functions of the class III PRX gene family.

2 Materials and methods

2.1 Identification of PRX family members

Genome data of R570 BAC and R570 STP were obtained from the sugarcane genome hub (https://sugarcane-genome.cirad.fr/; Garsmeur et al., 2018). Genome data of *A. thaliana* TAIR10, *Oryza sativa* (IRGSP-1.0), *S. bicolor* (NCBIv3), and *S. spontaneum* AP85-441 were obtained from the Ensembl Plants database (http://plants.ensembl.org/index.html; Yates et al., 2022).

The hidden Markov model (HMM) file of the PRX domain (PF00141) was downloaded from the Pfam database (https://pfam-legacy.xfam.org/; Sonnhammer et al., 1998; Blom et al., 1999; Mistry et al., 2021). The HMMER software (version 3.1b1; http://www.hmmer.org/) was used to search against the whole genome protein file of R570, *A. thaliana*, *O. sativa*, *S. bicolor*, and *S. spontaneum* under

the condition of e-value < 1×10^{-20}. Multiple sequence alignment was performed using ClustalW (http://www.clustal.org/), and a new HMM matrix file was constructed. The PRX domain-containing proteins were searched on the new HMM matrix file and screened by e-value < 0.001. The PRX protein Ref-seq of all plants was downloaded to build a library from the NCBI database (https://www.ncbi.nlm.nih.gov/). The putative PRX genes were searched using BLASTP and screened by e-value < 1×10^{-10} and identity > 75%. The protein-conserved domains were verified by Pfam and NCBI CDD (https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi), and the genes without the PRX domain were deleted (Lu et al., 2020; Mistry et al., 2021). In addition, the PRX gene family was identified based on the sugarcane transcriptome data.

The physicochemical properties of the *ShPRX* family proteins were calculated using ExPASy-ProtParam (https://web.expasy.org/protparam/; Wilkins et al., 1999). The subcellular localization of the *ShPRX* family proteins was predicted using Plant-mPLoc (http://www.csbio.sjtu.edu.cn/bioinf/plant-multi/; Chou and Shen, 2010). The transmembrane domain of the *ShPRX* family proteins was predicted using TMHMM 2.0 (https://services.healthtech.dtu.dk/service.php?TMHMM-2.0; Sonnhammer et al., 1998). The signal peptide of the *ShPRX* family proteins was predicted using SignalP 6.0 (https://services.healthtech.dtu.dk/service.php?SignalP; Teufel et al., 2022). The phosphorylation sites of the *ShPRX* family proteins were predicted using NetPhos 3.1 (https://services.healthtech.dtu.dk/service.php?NetPhos-3.1; Blom et al., 1999). The secondary structure of the *ShPRX* family proteins was analyzed using SOPMA (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html; Geourjon and Deleage, 1995).

2.2 Phylogenetic analysis

The multiple protein sequences of the PRX gene family in R570 STP, *A. thaliana*, *O. sativa*, *S. bicolor*, and *S. spontaneum* were compared using the MEGA-X MUSCLE software (https://www.megasoftware.net/; Kumar et al., 2018). The phylogenetic tree was constructed using the neighbor-joining method with the Jones–Taylor–Thornton model, 1,000 bootstrap replications, gamma distributed (G), and partial deletion gaps by MEGA-X. The phylogenetic tree was visualized using the online tool iTOL (https://itol.embl.de/; Letunic and Bork, 2021).

2.3 Chromosomal distribution, gene structure, and conserved motif analysis of the *ShPRXs*

The positional and structural information of *ShPRX* genes on the R570 STP chromosomes was extracted from the Generic Feature Format Version 3 (GFF3) (Garsmeur et al., 2018). The chromosomal locations of the *ShPRX* family genes were drawn using the online tool MG2C v2.1 (http://mg2c.iask.in/mg2c_v2.1/index.html). The conserved motifs of the *ShPRX* family proteins were identified using the online MEME with optimized parameters: minimum width, 6; maximum width, 50; and the number of motifs, 26 (Bailey and Elkan, 1994). The phylogenetic tree, gene structure, and conserved motifs of the *ShPRX* family genes were visualized using TBtools (Chen et al., 2020).

2.4 Analysis of *cis*-acting elements and codon usage bias of the *ShPRXs*

The sequence 2000 bp upstream from the start codon of *the ShPRX* family gene was obtained using the hybrid cultivar R570 reference genome (Garsmeur et al., 2018). The *cis*-acting elements of *ShPRX* family genes in promotor regions were predicted using the PlantCAR (Lescot et al., 2002). The phylogenetic tree and *cis*-acting elements of the *ShPRX* family genes were visualized using GSDS 2.0 (http://gsds.cbi.pku.edu.cn; Hu et al., 2015). The codon usage bias of the *ShPRX* family genes was analyzed using CodonW (https://codonw.sourceforge.net/) and EMBOSS: chips (https://www.bioinformatics.nl/cgi-bin/emboss/chips; Rice et al., 2000). The effective number of codon (ENc) plots and Parity rule 2 (PR2) plot analysis was performed as described by Wright (1990) and Chakraborty et al. (2020).

2.5 Syntenic and selection pressure analysis of the *PRX* family genes

The syntenic relationships of the PRX family genes were analyzed and visualized using MCScanX and Circos, respectively (Krzywinski et al., 2009; Wang et al., 2012). The frequencies of synonymous (Ks) and nonsynonymous (Ka) mutations, along with their ratios, were calculated to analyze the selection pressure of PRX duplicated gene pairs using TBtools (Chen et al., 2020). The divergence time (T) of class III PRX family genes was calculated as $T = Ks/(2 \times \lambda) \times 10^{-6}$ Mya ($\lambda = 6.5 \times 10^{-9}$ for grasses; (Gaut et al., 1996). The divergence time values were estimated using TimeTree (http://www.timetree.org/; Kumar et al., 2022).

2.6 Expression profiles of the *PRX* family genes

The RNA-seq expression data of the growth and development of *S. spontaneum* were downloaded from the Saccharum Genome Database (SGD; http://sugarcane.zhangjisenlab.cn/sgd/html/index.html; Zhang et al., 2018). The transcriptome data of sugarcane were obtained by sequencing (Akbar et al., 2021). The gene expression data of rice under Cd (GSE35502) and salt (GSE60287) stresses were downloaded from the GEO database (https://www.ncbi.nlm.nih.gov/geo/; Ishimaru et al., 2012; Garg et al., 2015; Clough and Barrett, 2016; Shankar et al., 2016). The expression pattern of the PRX family genes was drawn based on FPKM values or $\log_2$ fold change ($\log_2$FC) using the R software (v4.0.5).

2.7 SCMV infection, Cd treatment, salt stress treatment, and qRT-PCR analysis

Badila and B48, grown to leaf stages 6–8 in the field, was inoculated using SCMV crude extract following Yao's method (Yao et al., 2017), and the −3 and +1 leaves were collected at day 21 post-inoculation. The control plants were rubbed using 0.1 M phosphate buffer (pH 7.0). Sugarcane plants of Zhongzhe 1 were watered using 4.3-mM Cd solution or 0.5-L 1.0% sodium chloride solution when Zhongzhe 1 grew to leaf stage 3 in barrels containing 16 kilograms of soil. The control plants were watered using 0.5-L double-distilled water. The +1 leaves of sugarcane were collected at different time after treatment (0, 4, 8, 12, and 24 h).

Total RNA was extracted using Eastep® Super Total RNA Extraction Kit (Promega (Beijing) Biotech Co. Ltd., Beijing, China) following the manufacturer's instructions. The first strand of cDNA was synthesized using PrimeScript™ II 1st Strand cDNA Synthesis Kit (Takara Biomedical Technology Co. Ltd., Beijing, China). The qRT-PCR experiment was conducted in three independent replicates using ChamQ Universal SYBR qPCR Master Mix (Nanjing Vazyme Biotech Co., Ltd., Nanjing, China) with the Bio-Rad CFX96 fluorescence quantitative PCR Instrument. The primers for qRT-PCR of 10 *PRX* genes were designed using Primer Premier 6. The relative expression levels of the PRX family genes were calculated using the $2^{-\Delta\Delta Ct}$ method, and statistical significance was analyzed using ordinary one-way ANOVA in GraphPad Prism 7.

3 Results

3.1 Identification and chromosomal distribution of the *PRX* gene family in sugarcane

Eighty-two PRX proteins with a conserved PRX domain were characterized as members of the class III *PRX* gene family in R570 STP and named *ShPRX1–ShPRX82* based on their respective locations on the chromosomes (Tables S1 , S2). The total number of *PRX* family genes in the sugarcane R570 cultivar was lower than that in the *S. spontaneum* (113), sorghum (150), and rice (126), but slightly higher *in A. thaliana* (80) (Table S2).

The physical and chemical properties analysis revealed that the *ShPRX* family genes were predicted to encode polypeptides from 235 to 472 amino acids, with predicted molecular weights ranging from 25.11 to 49.76 kD. The theoretical pI ranged from 4.59 to 10.08, and the grand average of the hydropathicity values of 52 *ShPRX* proteins was negative, ranging from −0.51 to −0.002, indicating a hydrophilic characteristic. The grand average of the hydropathicity values of 30 *ShPRX* proteins was positive, ranging from 0.004 to 0.297, indicating a hydrophobic characteristic. The predicted number of negatively charged residues (Asp + Glu) in the *ShPRXs* was 14–59, and the number of positively charged residues (Arg + Lys) was 20–54. The instability index of 46 *ShPRX* proteins was less than 40, ranging from 25.38 to 39.76, indicating a stable characteristic. The instability index of 36 *ShPRX* proteins was greater than 40, ranging from 40.36 to 56.05, indicating an unstable characteristic (Table S1).

The subcellular localization prediction analysis showed that most *ShPRX* genes were located in the cytoplasm, and some were located in the vacuole, chloroplast, peroxisome, mitochondrion, cell membrane, and nucleus. Most *ShPRX* genes had signal peptides, 42 *ShPRX* genes had no transmembrane domain, and other *ShPRX* genes had only one transmembrane domain. The number of phosphorylation sites ranged from 16 to 47, while four *ShPRX* genes had no tyrosine phosphorylation site. The secondary structure prediction analysis showed that *ShPRX* proteins were mainly composed of alpha helix and random coil (Table S1).

The 82 *ShPRXs* were unevenly mapped onto the eight chromosomes of R570 STP (see Figure 1) based on the annotation information of the R570 STP genome. The *ShPRX* genes distributed in Sh01 (16), Sh02 (12), Sh03 (15), Sh04 (11), Sh06 (7), and Sh09 (10), while chromosomes Sh07 and Sh08 had only 4 and 3 *ShPRX* genes, respectively. Four *ShPRX* genes were distributed on scaffolds (Sh_011C11, Sh_025L09, and Sh_232D06).

3.2 Phylogenetic of the *PRX* gene family

The phylogenetic tree from five plants was constructed on PRX amino acid to clarify the evolutionary relationship of the *PRX* gene family of sugarcane, as shown in Figure 2. The *ShPRX* family genes were divided into six groups. The phylogenetic tree based on the amino acid sequences of *ShPRX* indicated that the *ShPRX* family genes in groups 1, 2, 3, 4, 5, and 6 from five plants were clustered into groups I, II, VI, V, III and IV from sugarcane, besides *ShPRX1* (group 3), *ShPRX34* (group 3), *ShPRX22* (group 4), and *ShPRX62* (group 2). *ShPRX1* and *ShPRX34* were divided into group V, and *ShPRX22* and *ShPRX62* were divided into group IV (Figure 2; Figure S1). Group 1 comprised the lowest number of *PRX* genes. The *PRX* family genes of sugarcane shared high homology with the *PRX* family genes in sorghum and rice (Figure 2). The *ShPRX* family genes in other groups shared high homology to *S. spontaneum* or sorghum (Figure 2).

3.3 Gene structure and conserved motif analyses of the *ShPRX* gene family

The conservative structure of the *PRX* gene family was deciphered through the evolutionary relationship, motif, and structure of the PRX family genes in sugarcane (Figure 3). MEME was used to analyze the motif distribution within the *ShPRX* gene family, and 26 motifs were identified ($p < 0.05$) (Table S3). The *ShPRX* genes in the same group had the same motifs: the genes in Group I had motifs 3, 8, 18, and 19; Group II had motifs 3, 4, 5, 8, 10, and 12; Group III had motifs 3, 8, and 11; Group IV had motifs 2, 4, 5, 6, and 7; group V had motifs 1, 4, 7, and 8; and Group VI had motif 3, and most *ShPRX* genes have top 14 motifs (Table S4; Figures 3A, B). The *ShPRX* family genes had motif 3 except for *ShPRX10* and *ShPRX43*, and most genes had motif 8 except for *ShPRX43* and *ShPRX44* (Figure 3B).

According to the structural analysis of the *PRX* family genes, the length of these genes ranged from 723 to 7812 bp, where *ShPRX67* had the most considerable length and 11 coding regions, *ShPRX47* had the most petite length (Table S1; Figure 3C, S2). The respective

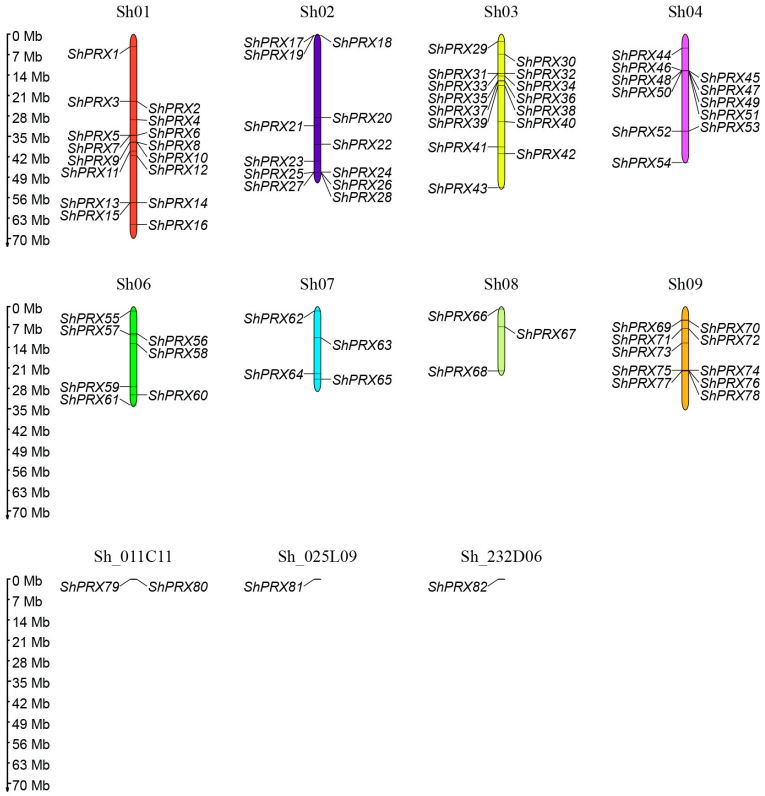


FIGURE 1
Distribution of the PRX gene family members on chromosomes in sugarcane.

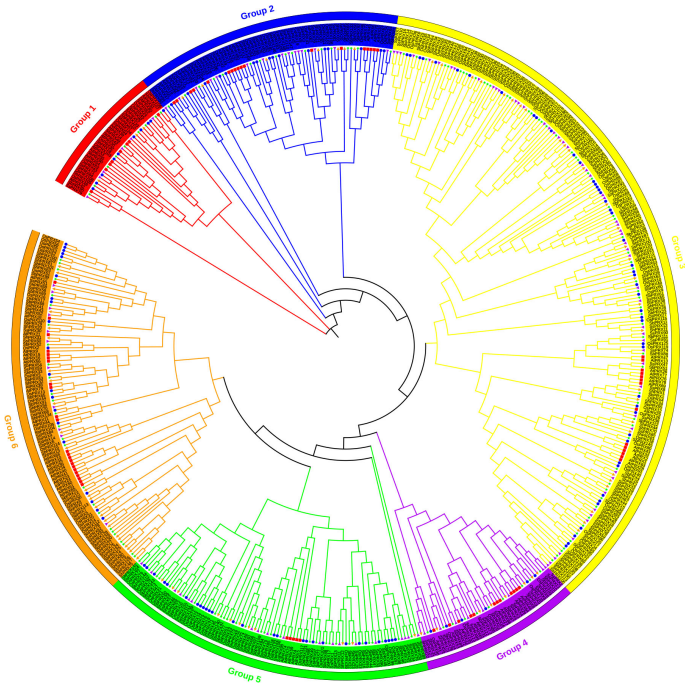


FIGURE 2
Phylogenetic analysis of PRX proteins from *A. thaliana*, *O. sativa*, *S. bicolor*, *S. spontaneum*, and *sugarcane*. *AtPRX*, red rectangle in the figure; *OsPRX*, blue circle in the figure; *SbPRX*, green star in the figure; *SsPRX*, purple right-pointing triangle in the figure; *ShPRX*, orange left-pointing triangle in the figure.

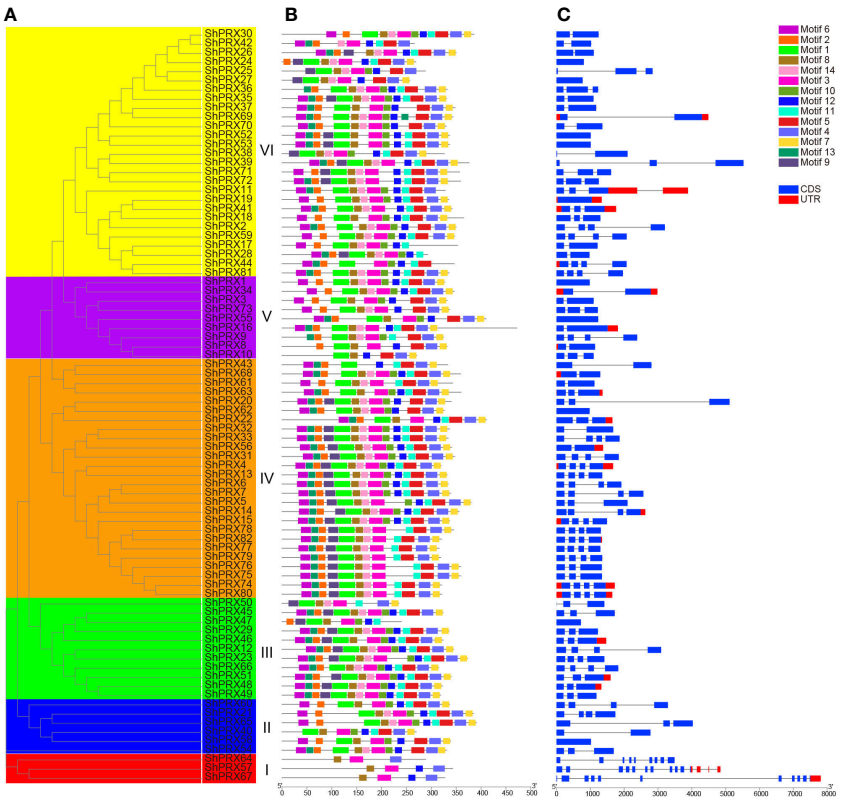


FIGURE 3
Analysis of gene structure and conserved motif of the PRX gene family in each group of sugarcane. **(A)** Phylogenetic tree of the *ShPRX* proteins. **(B)** Motif composition in the *ShPRX* proteins. **(C)** Gene structure of the *ShPRXs*.

structures of these genes were found to be similar in the same group, and 10 *PRX* genes had no introns. Most *PRX* genes contained 1–3 introns, but *ShPRX57*, *ShPRX64*, and *ShPRX67* in Group I contained 13, 8, and 10 introns, respectively, and *ShPRX20* (Group IV), *ShPRX39* (Group VI), and *ShPRX69* (Group VI) had long introns (Figures 3A, C).

3.4 *Cis*-acting elements and codon usage bias of *PRX* gene family in sugarcane

Cis-acting elements can participate in gene expression and regulation, and members of the *ShPRX* gene family play a key role in biotic and abiotic stress. The function of the *ShPRX* family genes can be predicted by analyzing the *cis*-acting elements in the 2-kb upstream region of the *ShPRX* family genes. Thirty-one cis-acting elements were identified and involved in hormones, abiotic stress, tissue-specific cell cycle, and circadian control by analyzing and selecting *cis*-acting elements of the *ShPRX* family genes (Table S5). The differences were detected in the variety and number of *cis*-acting elements across the *ShPRX* family genes (Table S5). *ShPRX15*, *ShPRX41*, *ShPRX48*, and *ShPRX62* had the maximum variety of *cis*-acting elements (15), but the number of cis-acting elements in *ShPRX52* was the greatest (37). *ShPRX22*, *ShPRX38*, *ShPRX49*, and *ShPRX62* contained plant hormone-responsive elements, including IAA-, GA-, ABA-, SA-, and MeJA-responsive elements. We inferred that plant hormones might regulate these *ShPRX* family genes. *ShPRX31* and *ShPRX66* did not contain ABA and light-responsive elements but contained seed-specific regulation elements and MYB binding sites involved in drought inducibility (Table S5). *The cis*-acting regulatory elements of most *ShPRX* family genes were involved in ABA, MeJA, light responsiveness, anaerobic induction, and drought inducibility (Figure 4A).

The cis-acting elements in Groups III, IV, V, and VI, were involved in the meristem-specific activation and expression. Over half of those in the Group I, II, and V were involved in the defense and stress responsiveness (Figures 4B, C). The *cis*-acting regulatory elements in Group 1 involved in the IAA, ABA, SA, MeJA, light, low-temperature, and anaerobic induction, and group IV in the seed-specific regulation (Table S6). These results suggested that the *ShPRX* family genes might participate in the response of sugarcane to biotic and abiotic stresses and tissue-specific responses.

The codon usage bias and base composition analysis of coding sequences of the *PRX* family genes were calculated (Table S7, S8). The mean values of the codon composition at the third position from high to low were C3s (0.589), G3s (0.440), T3s (0.093), and A3s (0.080), and the mean content of the GC (65.1%) was also higher than AT (34.9%), suggesting a GC-rich composition of coding sequences of the *PRX* family genes (Table S7). The ENC values of coding sequences ranged from 28.25 to 58.62, with a mean of 38.059 (ENC < 40), and most ENC values of coding sequences were below 40, indicating a solid codon usage bias (Table S7). The RSCU revealed that 22 codons of 28 high-frequency codons (mean RSCU value > 1.0) were over-represented (mean RSCU value > 1.6). In comparison, 31 codons of

10.3389/fpls.2023.1101665

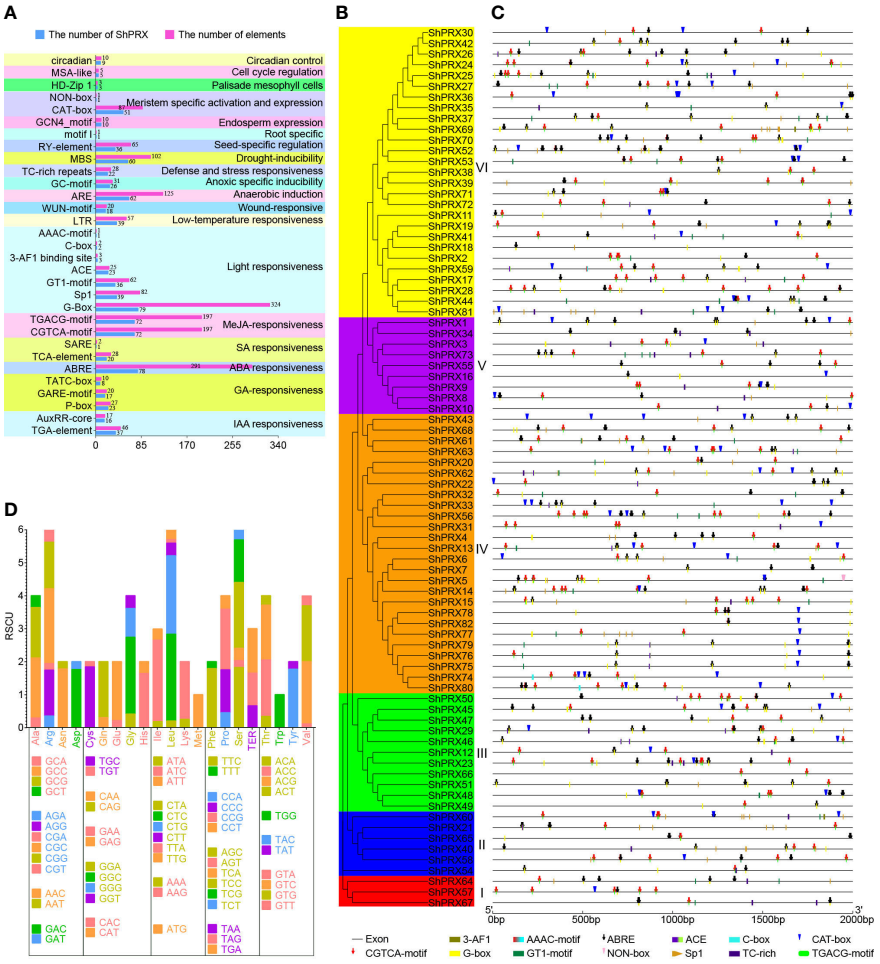


FIGURE 4
Analysis of cis-acting elements and codon usage bias of the PRX gene family in sugarcane. **(A)** The analysis of cis-acting elements of the PRX gene family in sugarcane. **(B)** The phylogenetic tree of the *ShPRX* proteins. **(C)** The analysis of cis-acting elements involved in the ABA, MeJA, light, defense, and stress responsiveness, meristem-specific activation, and expression in the PRX gene family of each group. **(D)** The codon usage bias analysis of PRX gene family in sugarcane.

34 low-frequency codons (mean RSCU value < 1.0) were under-represented (mean RSCU value < 0.6) (Figure 4D; Table S8). The ENc plot (Figure S3A) and PR2 plot (Figure S3B) analyses showed that the codon use of the class III PRX family genes in sugarcane was affected by mutation and selection pressure.

3.5 Syntenic and selection pressure analysis of the *ShPRX* gene family

The syntenic relationships of the *ShPRXs* were analyzed to explore the genomic expansion of the *PRX* gene family in sugarcane. In total, 32 of the 82 *ShPRXs* had syntenic relationships, and 10 ones in five syntenic pairs underwent segmental duplication, while 22 genes in 13 syntenic pairs underwent tandem duplication (Figure 5A; Table S9). The Ka/Ks ratios of 17 of the 18 syntenic pairs were < 1, which might have undergone purifying selection, indicating that the evolution of these pairs was slow (Figure 5A; Table S9).

The syntenic relationships of the PRX family genes from *S officinarum*, *A. thaliana*, *O. sativa*, *S. bicolor*, and *S. spontaneum* were analyzed to better understand duplication and evolution of *PRX* gene. The results revealed that 52, 68, and 37 syntenic gene pairs of PRX genes were detected in *S officinarum versus O. sativa*, *S officinarum versus S. bicolor*, and *S officinarum versus S. spontaneum*, respectively (Figures 5B, C; Table S9). The Ka/Ks ratio of 1 of the 157 syntenic gene pairs was> 1, indicating that the syntenic gene pair had been strongly positively selected during evolutionary history (Table S9). The Ka/Ks ratios of 153 from the 157 syntenic gene pairs were < 1, which might have undergone purifying selection, indicating that the evolution of these pairs was slow. The duplication events of homologous collinearity gene pairs of the class III PRX occurred 4.094–85.36 Mya for sugarcane and rice, 21.992–83.6 Mya for sugarcane and sorghum, and 3.568–50.095 Mya for sugarcane and its wild relative of *S. spontaneum*. The class III PRX gene family has been identified in at least 29 plants in the whole genome. The duplication events occurred 159.9 Mya for 12 plants in

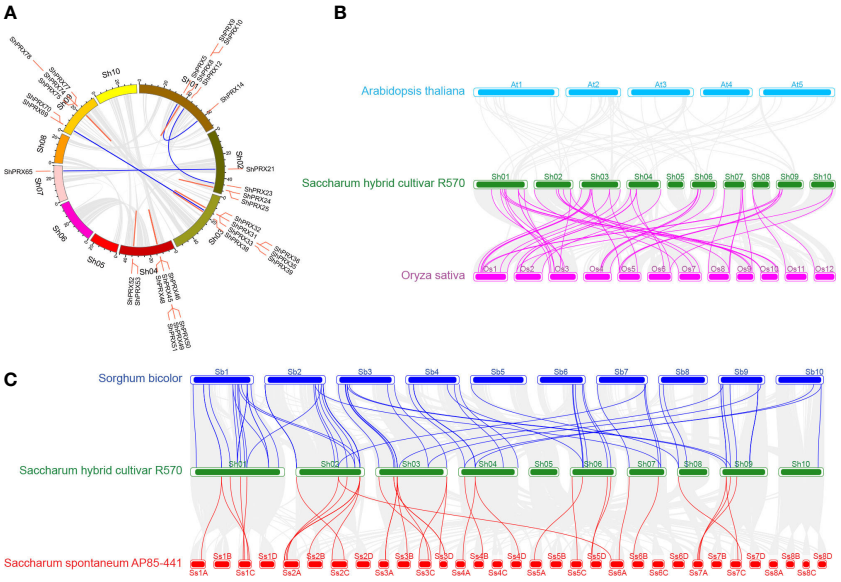


FIGURE 5
Syntenic analysis of PRX gene family in Sugarcane, *A. thaliana*, *O. sativa*, *S. bicolor*, and *S. spontaneum*. **(A)** Syntenic analysis of PRX gene family in sugarcane. Red lines represent tandem duplication PRX gene pairs, and blue lines represent segmental duplication PRX gene pairs. **(B)** Syntenic analysis of PRX gene family in Sugarcane, *A. thaliana*, and *O. sativa*. **(C)** Syntenic analysis of the PRX gene family in Sugarcane, *S. bicolor*, and *S. spontaneum*.

monocotyledon from 21 plants in dicotyledon. Moreover, 104.7 Mya in *Poaceae* diverged from *Bromeliaceae* (Figure S3C; Table S10).

3.6 Expression profiles of the *PRX* family genes in different tissue

Tissue-specific expression patterns of the *SsPRX* genes were analyzed based on the transcriptome data of *S. spontaneum* to explore the functions of the *PRX* gene family in sugarcane further. The 113 *SsPRX* genes were expressed as FPKM values in leaf roll, leaf, the third, sixth, and ninth stem nodes of sugarcane at the seedling, elongation and maturity stage (Figure 6A; Table S11). Four *SsPRX* genes (*SsPRX42*, *SsPRX62*, *SsPRX79*, and *SsPRX83*) were highly expressed in all tissues, and 69 *SsPRX* genes had a low expression or no expression, and the rest of the highly expressed *SsPRX* genes had tissue specificity.

3.7 Expression profiles of *PRX* family genes under SCMV, Cd, and salt stresses

Sixteen PRX proteins with a conserved PRX domain were characterized as members of the sugarcane class III *PRX* gene family based on the transcriptome data of Badila and B48 in response to SCMV infection. We named them *ShtPRX1–ShtPRX16* based on their Unigene IDs (Table S2). The expression patterns of the *ShtPRX* genes were analyzed on transcriptome data (Table S12). The *ShtPRX* genes were divided into two groups according to the clustering of expression patterns. Genes in Group I was primarily highly expressed in all samples. Genes in Group II, except for *ShtPRX13* and *ShtPRX16*, were primarily highly expressed in some samples (Figure 6B). ShtPRX8 was only one co-differentially expressed gene ($|\log_2 FC| \geq 3$, FDR < 0.05) in five groups (Figure 6C), indicating that *ShtPRX8* might participate in SCMV stress in sugarcane.

Homologous class III *PRX* genes of R570 were identified using class III PRX family protein sequences of rice as a library, and screened bu BLASTP at the e-value less than $1e^{-5}$. (Table S13). The expression profiles of rice under Cd and salt stresses were analyzed to infer the function of *PRX* homologous genes in sugarcane (Tables S14, S15; Figure S4). In groups #I, III, and VII, except for *OsPRX10*, the $|\log_2 FC|$ values for (Cd/CK)-WR or (Cd/CK)-WS was more than 3, and homologous genes, including *ShPRX38*, *ShBAC.PRX23*, *ShBAC.PRX41*, *ShBAC.PRX42*, *ShBAC.PRX43*, *ShBAC.PRX51*, *ShBAC.PRX52*, and *ShBAC.PRX75* might participate in Cd stress in sugarcane (Figure S4A; Table S13). In group 2, the $|\log_2 FC|$ of *OsPRX56* for PK-SS/PK-CK was more than 2, and homologous genes of *ShBAC.PRX20* might participate in salt stress in Sugarcane (Figure S4B; Tables S13, S14).

qRT-PCR analysis was performed to analyze the responses of the *PRX* family genes in sugarcane exposed to SCMV, Cd, and salt stresses. The primer sequences were listed in Table S16. The expression levels of *ShtPRX8* showed a significant decrease in Badila after SCMV infection. After applying 4.3 mM Cd^{2+} stress to sugarcane, the expression levels of *ShBAC.PRX23*, *ShBAC.PRX41*, *ShBAC.PRX42*, *ShBAC.PRX43*, *ShBAC.PRX51*, *ShBAC.PRX52*, *ShBAC.PRX75* and *ShPRX38* showed a significant increase from 0 to 24 h (Figure 7). After applying 1% sodium chloride solution stress to sugarcane, the expression levels of *ShBAC PRX20* significantly increased from 0 to 24 h (Figure 7).

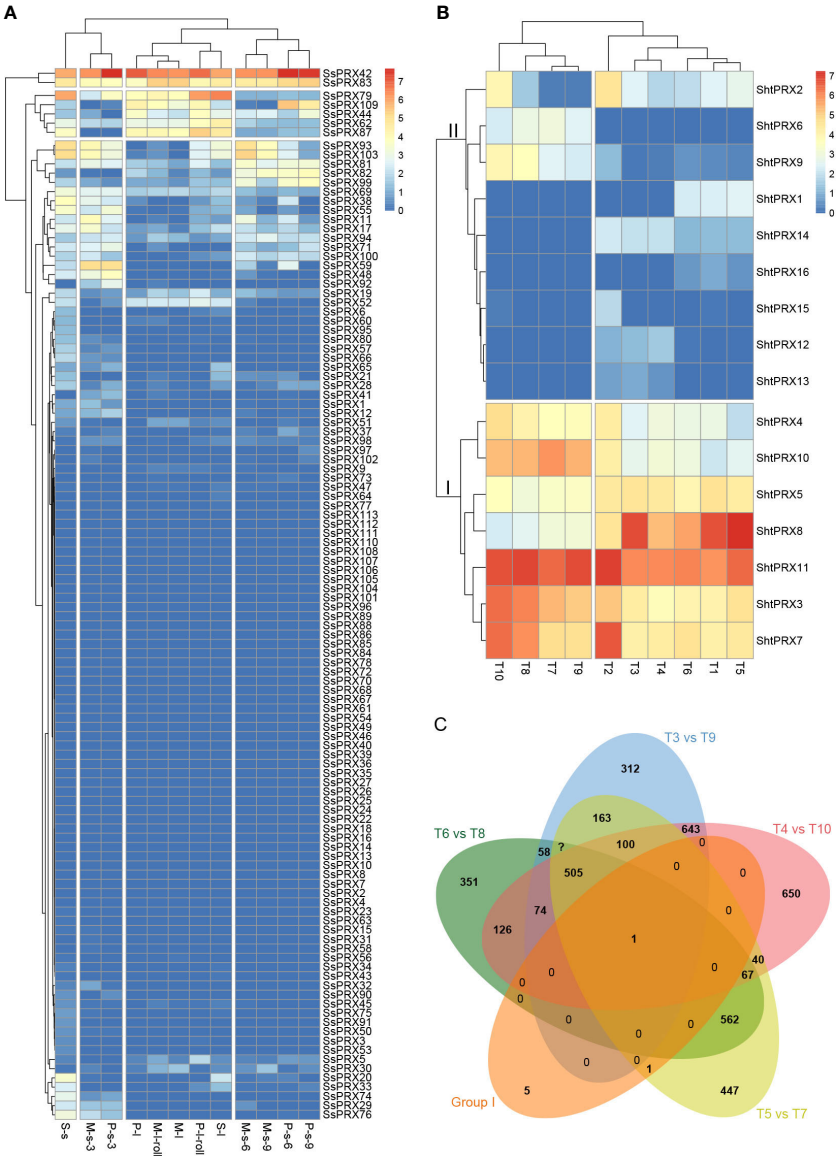


FIGURE 6
Expression pattern of the PRX genes in sugarcane. **(A)** Tissue-specific expression patterns of PRX gene family in *S. spontaneum*. S represents seedling; s represents stem; l represents leaf; P represents early maturity stage; M represents mature stage; 3, 6, and 9 represent sugarcane's third, sixth, and ninth stem nodes. **(B)** Expression patterns of PRX gene family of sugarcane in response to SCMV infection. T1, the +1 leaf of Badila; T2, the −3 leaf of Badila; T3, the +1 leaf of virus-free Badila; T4, the −3 leaf of virus-free Badila; T5, the +1 leaf of B48; T6, the −3 leaf of B48; T7, the +1 leaf of B48 post-SCMV infection; T8, the −3 leaf of B48 post-SCMV infection, T9, the +1 leaf of virus-free Badila post-SCMV infection; T10, the −3 leaf of virus-free Badila post-SCMV infection. **(C)** The number of co-differentially expressed genes in groups I, T6_vs_T8, T3_vs_T9, T4_vs_T10, and T5_vs_T7.

4 Discussion

Class III PRXs are essential in response to various biotic and abiotic stresses during plant growth and development. The class III *PRX* gene family has been identified in at least 29 plants in the whole genome, but the class III *PRX* gene family has yet to be studied in sugarcane. The number of *PRX* family genes identified in R570 BAC and STP was fewer than in other monocotyledons and dicotyledons. The class III *PRX* family genes were divided into six groups based on the phylogenetic analysis of sugarcane, sorghum, *S. spontaneum*, rice, and *A. thaliana* and the structure and conservative motif of class III PRX genes were similar in each group. The class III *PRX* genes in the same group were highly conserved. The identified class III *PRXs* in sugarcane had no collinear gene pair in *A. thaliana*. This is the same as Cesarino's results, showing three sugarcane *PRXs*, being monocot-specific sequences, had no clear ortholog in *A. thaliana* (Cesarino et al., 2012).

Poaceae diverged from *Bromeliaceae* at 104.7 Mya. The duplication events of homologous collinearity gene pairs of the class III PRX for sugarcane and rice, sugarcane and sorghum, and sugarcane and its wild relative of *S. spontaneum* occurred at 0–85.36 Mya. These results revealed that class III *PRXs* in sugarcane was formed after *Poaceae* and *Bromeliaceae* diverged. Tandem duplication events play a leading role in the expansion of the class III *PRX* gene family in sugarcane, which was the same as *B. distachyon* (Zhu et al., 2019), foxtail millet (Ma et al., 2022), and grapevine (Xiao et al., 2020).

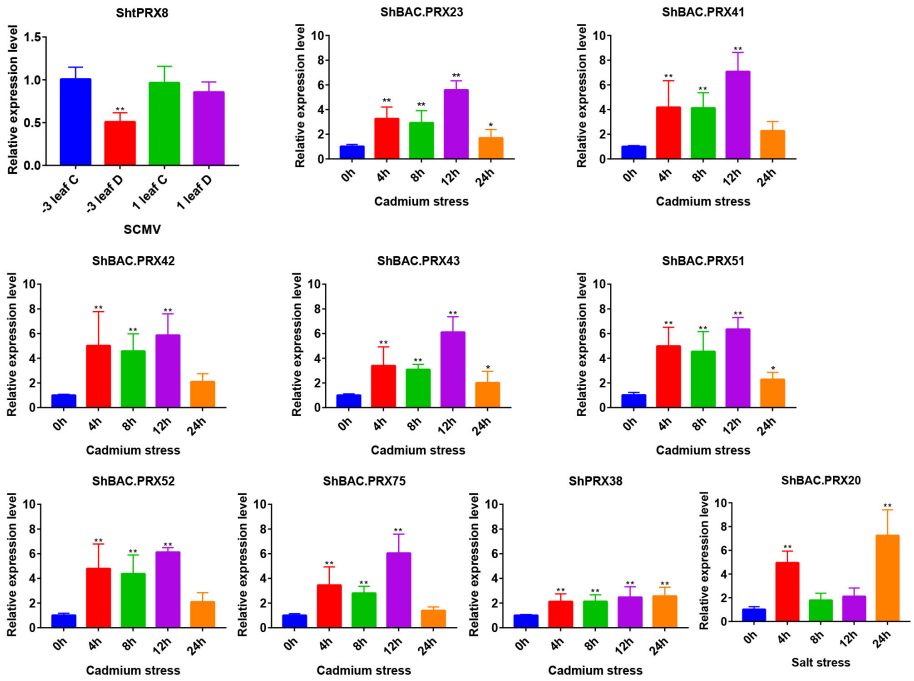


FIGURE 7
Analysis of SCMV, Cd, and salt stress differentially expressed PRX gene family members in sugarcane. The black bar graphs represent the relative expression levels of PRX family genes in the leaves under SCMV, Cd, and salt stresses. The −3 leaf C, the −3 leaf of virus-free Badila; the 1 leaf C, the +1 leaf of virus-free Badila; −3 leaf D, the -3 leaf of post-SCMV infection Badila (symptomatic leaf); 1 leaf D, the +1 leaf of post-SCMV infection Badila (asymptomatic leaf); *: 0.01 < p < 0.05, **: p < 0.01; The mean and SD were calculated from three biological and three technical replicate samples.

Purifying selection essentially maintained the function of class III PRX proteins in sugarcane. However, tandem and segmental duplications are the primary reasons for expanding the class III *PRX* gene family in maize (Wang et al., 2015) and pineapple (Hou et al., 2022). Segmental duplication is the main reason for the expansion of the class III PRX gene family in tobacco (Cheng et al., 2022), soybean (Aleem et al., 2022), and potato (Yang et al., 2020). These results indicated that tandem and segmental duplications were pivotal in expanding plants' class III *PRX* gene family.

Phylogenetic analysis revealed that *ShPRX* family genes shared high homology with the *PRX* family genes in *S. spontaneum* and sorghum, which was consistent with the genetic relationship between species. We inferred that *ShPRX* family genes might have the same function in the same group. GO enrichment analysis of motifs using InterProScan revealed that motifs 1, 2, 3, 4, 5, 18, 19, and 21 are essential in response to oxidative stress (GO:0006979) and have peroxidase activity (GO:0004601). However, there are some motifs with unknown functions in each group; these motifs might play a key role in the *ShPRX* family genes.

Gene expression patterns and cis-acting elements can provide important information regarding gene function. More than 87% of class III *ShPRX* genes were involved in light (G-box), ABA (ABRE), and MeJA (CGTCA and TGACG motifs) response, and more than 73% of class III PRX genes participated in anaerobic (ARE) and drought (MBS) responses. Also, over 62% of class III *PRX* genes participated in meristem-specific (CAT box) responses (Table S5). We inferred that class III *PRX* genes of sugarcane played an essential role in plant growth and development. The highest enzymatic activity was presented in the pith and rind of mature internodes at three different developmental stages of sugarcane (young, developing, and mature) (Cesarino et al., 2012). The expression of class III PRX proteins in susceptible sugarcane was inhibited for 72 h after *Sporisorium scitamineum* inoculation (Peters et al., 2017). Of 113 *SsPRX* genes, 69 exhibited little or no expression in tissues, indicating that *SsPRX* genes might be expressed under specific conditions or at specific developmental stages. Of 44 highly expressed genes, 26 and 43 were expressed in leaves and stems, respectively, and most of them might play an essential role in the leaf and stem of sugarcane.

In rice, the overexpression of *OsPrx30* contributed to maintaining a high level of PRX activity and reducing H_2O_2 content, thereby enhancing the rice plant's susceptibility to *Xoo* (Liu et al., 2021), suggesting that class III *PRX* genes might have similar functions in sugarcane. The concentrations of H_2O_2 increased significantly in B48. The genes related to ROS-producing and scavenging pathways were differentially expressed on the ScMV-inoculated plants at days 3, 6, 9, and 12 (Akbar et al., 2020). The expression of *ShtPRX8* was significantly reduced post-SCMV inoculation based on the transcriptional data from our previous study. PRX, primarily existing in peroxisomes, could reduce the ROS level. SCMV could target intracellular peroxisomes for replication (Xie et al., 2021). The class III *PRX* family genes might play an essential role in sugarcane mosaic disease by activating the antioxidant system and regulating ROS and H_2O_2 content. Plants produced ROS under salt and Cd stresses, and the class III *PRX* family genes play an essential role in plant Cd and salt stress by scavenging ROS (Chiang et al., 2006; Kidwai et al., 2020; Su et al., 2020). The overexpression of the class III *PRX* family gene of *TaPRX-2A* in wheat activated the ABA pathway

and antioxidant enzymes, leading to reduced ROS accumulation and increased osmotic metabolites, thereby enhancing salt tolerance (Su et al., 2020). The expression levels of class III PRXs significantly increased after applying 4.3 mM Cd and 1% NaCl solution stress to sugarcane. We inferred that the class III PRX family genes could enhance the tolerance of Cd and salt stresses in sugarcane by activating the antioxidant system and scavenging ROS.

5 Conclusion

Tandem duplication events play a leading role in the expansion of *ShPRX* genes, and purifying selection essentially maintains the function of *ShPRX* proteins. In this study, 82 *ShPRX* genes were identified in the R570 STP genome and divided them into six groups. Expression profile and qRT-PCR analyses showed that SCMV, Cd, and salt could specifically induce the expression of *PRX* genes of sugarcane. These results help understand the structure, evolution, and functions of the class III PRX gene family in sugarcane with a view of providing ideas for the phytoremediation of Cd-contaminated soil and breeding of new sugarcane varieties resistant to sugarcane mosaic disease, salt, and Cd in the future.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

Author contributions

MZ, WY, and JX designed the research. HS wrote the draft manuscript. LF, LQ, HJ, ZD, HZ, and ZY performed the experiments and data analyses. GC and YB conducted the genomic analysis. JX and WY edited the first draft. MZ conceived the idea, provided supervision, revised the manuscript, and provided funds for this investigation. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by the National Natural Science Foundation of China (32001603), the Science and Technology Major Project of Guangxi (Gui Ke AA22117007, Gui Ke AB21238008), and the China Agriculture Research System of MOF and MARA (CARS-170109, CARS-170726). The authors would like to thank the reviewers for their helpful comments on this manuscript.

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1101665/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Phylogenetic analysis of ShPRX proteins.

SUPPLEMENTARY FIGURE 2
Analysis of gene structure and conserved motif of the PRX gene family.

References

Akbar, S., Wei, Y., Yuan, Y., Khan, M. T., Qin, L., et al. (2020). Gene expression profiling of reactive oxygen species (ROS) and antioxidant defense system following sugarcane mosaic virus (SCMV) infection. *BMC Plant Biol.* 20, 532. doi: 10.1186/s12870-020-02737-1

Akbar, S., Yao, W., Qin, L., Yuan, Y., Powell, C. A., et al. (2021). Comparative analysis of sugar metabolites and their transporters in sugarcane following sugarcane mosaic virus (SCMV) infection. *Int. J. Mol. Sci.* 22. doi: 10.3390/ijms222413574

Aleem, M., Riaz, A., Raza, Q., Aleem, M., Aslam, M., et al. (2022). Genomewide characterization and functional analysis of class iii peroxidase gene family in soybean reveal regulatory roles of gspod40 in drought tolerance. *Genomics* 114, 45–60. doi: 10.1016/j.ygeno.2021.11.016

Bailey, T. L., and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf Intell. Syst. Mol. Biol.* 2, 28–36.

Blom, N., Gammeltoft, S., and Brunak, S. (1999). Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J. Mol. Biol.* 294, 1351–1362. doi: 10.1006/jmbi.1999.3310

Cesarino, I., Araujo, P., Sampaio, M. J., Paes, L. A., and Mazzafera, P. (2012). Enzymatic activity and proteomic profile of class iii peroxidases during sugarcane stem development. *Plant Physiol. Biochem.* 55, 66–76. doi: 10.1016/j.plaphy.2012.03.014

Chakraborty, S., Yengkhom, S., and Uddin, A. (2020). Analysis of codon usage bias of chloroplast genes in oryza species: codon usage of chloroplast genes in oryza species. *Planta* 252, 67. doi: 10.1007/s00425-020-03470-7

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

Chen, Y., Feng, J., Qu, Y., Zhang, J., Zhang, L., et al. (2022). Genomewide identification and functional analysis of class iii peroxidases in gossypium hirsutum. *Peer J.* 10, e13635. doi: 10.7717/peerj.13635

Cheng, L., Ma, L., Meng, L., Shang, H., Cao, P., et al. (2022). Genomewide identification and analysis of the class iii peroxidase gene family in tobacco (Nicotiana tabacum). *Front. Genet.* 13. doi: 10.3389/fgene.2022.916867

Chiang, H. C., Lo, J. C., and Yeh, K. C. (2006). Genes associated with heavy metal tolerance and accumulation in Zn/Cd hyperaccumulator arabidopsis halleri: a genomic survey with cDNA microarray. *Environ. Sci. Technol.* 40, 6792–6798. doi: 10.1021/es061432y

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

Clough, E., and Barrett, T. (2016). The gene expression omnibus database. *Methods Mol. Biol.* 1418, 93–110. doi: 10.1007/978-1-4939-3578-9_5

Duan, P., Wang, G., Chao, M., Zhang, Z., and Zhang, B. (2019). Genomewide identification and analysis of class iii peroxidases in allotetraploid cotton (Gossypium hirsutum l.) and their responses to pk deficiency. *Genes (Basel)* 10. doi: 10.3390/genes10060473

Francoz, E., Ranocha, P., Nguyen-Kim, H., Jamet, E., Burlat, V., et al. (2015). Roles of cell wall peroxidases in plant development. *Phytochemistry* 112, 15–21. doi: 10.1016/j.phytochem.2014.07.020

Garg, R., Narayana, C. V., Shankar, R., and Jain, M. (2015). Divergent DNA methylation patterns associated with gene expression in rice cultivars with contrasting drought and salinity stress response. *Sci. Rep.* 5, 14922. doi: 10.1038/srep14922

Garsmeur, O., Droc, G., Antonise, R., Grimwood, J., Potier, B., et al. (2018). A mosaic monoploid reference sequence for the highly complex genome of sugarcane. *Nat. Commun.* 9, 2638. doi: 10.1038/s41467-018-05051-5

Gaut, B. S., Morton, B. R., McCaig, B. C., and Clegg, M. T. (1996). Substitution rate comparisons between grasses and palms: synonymous rate differences at the nuclear gene adh parallel rate differences at the plastid gene rbcl. *Proc. Natl. Acad. Sci. U S* 93, 10274–10279. doi: 10.1073/pnas.93.19.10274

Geourjon, C., and Deleage, G. (1995). Sopma: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Comput. Appl. Biosci.* 11, 681–684. doi: 10.1093/bioinformatics/11.6.681

Hou, X., Lu, Z., Hong, K., Song, K., Gu, H., et al. (2022). The class iii peroxidase gene family is involved in ascorbic acid induced delay of internal browning in pineapple. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.953623

Hu, B., Jin, J., Guo, A. Y., Zhang, H., Luo, J., et al. (2015). Gsds 2.0: an upgraded gene feature visualization server. *Bioinformatics* 31, 1296–1297. doi: 10.1093/bioinformatics/btu817

Ishimaru, Y., Takahashi, R., Bashir, K., Shimo, H., Senoura, T., et al. (2012). Characterizing the role of rice nramp5 in manganese, iron, and cadmium transport. *Sci. Rep.* 2, 286. doi: 10.1038/srep00286

Kidwai, M., Ahmad, I. Z., and Chakrabarty, D. (2020). Class iii peroxidase: an indispensable enzyme for biotic/abiotic stress tolerance and a potent candidate for crop improvement. *Plant Cell Rep.* 39, 1381–1393. doi: 10.1007/s00299-020-02588-y

Kidwai, M., Dhar, Y. V., Gautam, N., Tiwari, M., Ahmad, I. Z., et al. (2019). Oryza sativa class iii peroxidase (osprx38) overexpression in arabidopsis thaliana reduces arsenic accumulation due to apoplastic lignification. *J. Hazard Mater* 362, 383–393. doi: 10.1016/j.jhazmat.2018.09.029

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

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

Kumar, S., Suleski, M., Craig, J. M., Kasprowicz, A. E., Sanderford, M., et al. (2022). Timetree 5: an expanded resource for species divergence times. *Mol. Biol. Evol.* doi: 10.1093/molbev/msac174

Lee, T. T. (1977). Role of phenolic inhibitors in peroxidase-mediated degradation of indole-3-acetic acid. *Plant Physiol.* 59, 372–375. doi: 10.1104/pp.59.3.372

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

Letunic, I., and Bork, P. (2021). Interactive tree of life (itol) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 49, W293–W296. doi: 10.1093/nar/gkab301

Li, Q., Dou, W., Qi, J., Qin, X., Chen, S., et al. (2020). Genomewide analysis of the ciii peroxidase family in sweet orange (Citrus sinensis) and expression profiles induced by xanthomonas citri subsp. citri and hormones. *J. Genet.* 99. doi: 10.1007/s12041-019-1163-5

Liu, H., Dong, S., Li, M., Gu, F., Yang, G., et al. (2021). The class iii peroxidase gene osprx30, transcriptionally modulated by the at-hook protein osath1, mediates rice bacterial blight-induced ROS accumulation. *J. Integr. Plant Biol.* 63, 393–408. doi: 10.1111/jipb.13040

Lu, S., Wang, J., Chitsaz, F., Derbyshire, M. K., Geer, R. C., et al. (2020). Cdd/sparcle: the conserved domain database in 2020. *Nucleic Acids Res.* 48, D265–D268. doi: 10.1093/nar/gkz991

Ma, X., Xu, R., Suo, X., Li, J., Gu, P., et al. (2022). Genomewide identification of the class iii prx gene family in foxtail millet (Setaria italica l.) and expression analysis under drought stress. *Acta Agronomica Sin.* 10, 2517–2532.

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., et al. (2021). Pfam: the protein families database in 2021. *Nucleic Acids Res.* 49, D412–D419. doi: 10.1093/nar/gkaa913

Moradi, Z., Mehrvar, M., and Nazifi, E. (2018). Genetic diversity and biological characterization of sugarcane streak mosaic virus isolates from Iran. *Virus Dis.* 29, 316–323. doi: 10.1007/s13337-018-0461-5

Passardi, F., Longet, D., Penel, C., and Dunand, C. (2004). The class iii peroxidase multigenic family in rice and its evolution in land plants. *Phytochemistry* 65, 1879–1893. doi: 10.1016/j.phytochem.2004.06.023

Peters, L. P., Carvalho, G., Vilhena, M. B., Creste, S., Azevedo, R. A., et al. (2017). Functional analysis of oxidative burst in sugarcane smut-resistant and -susceptible genotypes. *Planta* 245, 749–764. doi: 10.1007/s00425-016-2642-z

Piontek, K., Smith, A. T., and Blodig, W. (2001). Lignin peroxidase structure and function. *Biochem. Soc. Trans.* 29, 111–116. doi: 10.1042/0300-5127:0290111

Rice, J. L., and Hoy, J. W. (2020). Recovery from mosaic caused by sorghum mosaic virus in sugarcane and impact on yield. *Plant Dis.* 104, 3166–3172. doi: 10.1094/PDIS-02-20-0376-RE

Rice, J. L., Hoy, J. W., and Grisham, M. P. (2019). Sugarcane mosaic distribution, incidence, increase, and spatial pattern in Louisiana. *Plant Dis.* 103, 2051–2056. doi: 10.1094/PDIS-01-19-0099-RE

Rice, P., Longden, I., and Bleasby, A. (2000). Emboss: the European molecular biology open software suite. *Trends Genet.* 16, 276–277. doi: 10.1016/s0168-9525(00)02024-2

Sereno, M. L., Almeida, R. S., Nishimura, D. S., and Figueira, A. (2007). Response of sugarcane to increasing concentrations of copper and cadmium and expression of metallothionein genes. *J. Plant Physiol.* 164, 1499–1515. doi: 10.1016/j.jplph.2006.09.007

Shankar, R., Bhattacharjee, A., and Jain, M. (2016). Transcriptome analysis in different rice cultivars provides novel insights into desiccation and salinity stress responses. *Sci. Rep.* 6, 23719. doi: 10.1038/srep23719

Shigeto, J., and Tsutsumi, Y. (2016). Diverse functions and reactions of class iii peroxidases. *New Phytol.* 209, 1395–1402. doi: 10.1111/nph.13738

Sonnhammer, E. L., von Heijne, G., and Krogh, A. (1998). A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc. Int. Conf Intell. Syst. Mol. Biol.* 6, 175–182.

Su, P., Yan, J., Li, W., Wang, L., Zhao, J., et al. (2020). A member of wheat class iii peroxidase gene family, taprx-2a, enhanced the tolerance of salt stress. *BMC Plant Biol.* 20, 392. doi: 10.1186/s12870-020-02602-1

Teufel, F., Almagro, A. J., Johansen, A. R., Gislason, M. H., Pihl, S. I., et al. (2022). Signalp 6.0 predicts all five types of signal peptides using protein language models. *Nat. Biotechnol.* 40, 1023–1025. doi: 10.1038/s41587-021-01156-3

Tognolli, M., Penel, C., Greppin, H., and Simon, P. (2002). Analysis and expression of the class iii peroxidase large gene family in arabidopsis thaliana. *Gene* 288, 129–138. doi: 10.1016/s0378-1119(02)00465-1

Vanholme, R., De Meester, B., Ralph, J., and Boerjan, W. (2019). Lignin biosynthesis and its integration into metabolism. *Curr. Opin. Biotechnol.* 56, 230–239. doi: 10.1016/j.copbio.2019.02.018

Wang, Y., Tang, H., Debarry, J. D., Tan, X., Li, J., et al. (2012). Mcscanx: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 40, e49. doi: 10.1093/nar/gkr1293

Wang, Y., Wang, Q., Zhao, Y., Han, G., and Zhu, S. (2015). Systematic analysis of maize class iii peroxidase gene family reveals a conserved subfamily involved in abiotic stress response. *Gene* 566, 95–108. doi: 10.1016/j.gene.2015.04.041

Wilkins, M. R., Gasteiger, E., Bairoch, A., Sanchez, J. C., Williams, K. L., et al. (1999). Protein identification and analysis tools in the expasy server. *Methods Mol. Biol.* 112, 531–552. doi: 10.1385/1-59259-584-7:531

Wright, F. (1990). The 'effective number of codons' used in a gene. *Gene* 87, 23–29. doi: 10.1016/0378-1119(90)90491-9

Xiao, H., Wang, C., Khan, N., Chen, M., Fu, W., et al. (2020). Genomewide identification of the class iii pod gene family and their expression profiling in grapevine (vitis vinifera l). *BMC Genomics* 21, 444. doi: 10.1186/s12864-020-06828-z

Xie, J., Jiang, T., Li, Z., Li, X., Fan, Z., et al. (2021). Sugarcane mosaic virus remodels multiple intracellular organelles to form genomic RNA replication sites. *Arch. Virol.* 166, 1921–1930. doi: 10.1007/s00705-021-05077-z

Yang, X., Yuan, J., Luo, W., Qin, M., Yang, J., et al. (2020). Genomewide identification and expression analysis of the class iii peroxidase gene family in potato (Solanum tuberosum l.). *Front. Genet.* 11. doi: 10.3389/fgene.2020.593577

Yao, W., Ruan, M., Qin, L., Yang, C., Chen, R., et al. (2017). Field performance of transgenic sugarcane lines resistant to sugarcane mosaic virus. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.00104

Yates, A. D., Allen, J., Amode, R. M., Azov, A. G., Barba, M., et al. (2022). Ensembl genomes 2022: an expanding genome resource for non-vertebrates. *Nucleic Acids Res.* 50, D996–D1003. doi: 10.1093/nar/gkab1007

Yousefi, Z., Kolahi, M., Majd, A., and Jonoubi, P. (2018). Effect of cadmium on morphometric traits, antioxidant enzyme activity, and phytochelatin synthase gene expression (sopcs) of saccharum officinarum var. Cp48-103 *in vitro*. *Ecotoxicol Environ. Saf.* 157, 472–481. doi: 10.1016/j.ecoenv.2018.03.076

Zamocky, M. (2004). Phylogenetic relationships in class I of the superfamily of bacterial, fungal, and plant peroxidases. *Eur. J. Biochem.* 271, 3297–3309. doi: 10.1111/j.1432-1033.2004.04262.x

Zamocky, M., Furtmuller, P. G., and Obinger, C. (2010). Evolution of structure and function of class I peroxidases. *Arch. Biochem. Biophys.* 500, 45–57. doi: 10.1016/j.abb.2010.03.024

Zhang, H., Wang, Z., Li, X., Gao, X., Dai, Z., et al. (2022). The ibbbx24-ibtoe3-ibprx17 module enhances abiotic stress tolerance by scavenging reactive oxygen species in sweet potato. *New Phytol.* 233, 1133–1152. doi: 10.1111/nph.17860

Zhang, J., Zhang, X., Tang, H., Zhang, Q., Hua, X., et al. (2018). Allele-defined genome of the autopolyploid sugarcane saccharum spontaneum l. *Nat. Genet.* 50, 1565–1573. doi: 10.1038/s41588-018-0237-2

Zhu, X., Jiang, L., Cai, Y., and Cao, Y. (2021). Functional analysis of four class iii peroxidases from Chinese pear fruit: a critical role in lignin polymerization. *Physiol. Mol. Biol. Plants* 27, 515–522. doi: 10.1007/s12298-021-00949-9

Zhu, T., Li, L., Duan, Q., Liu, X., and Chen, M. (2021). Progress in our understanding of plant responses to the stress of heavy metal cadmium. *Plant Signal Behav.* 16, 1836884. doi: 10.1080/15592324.2020.1836884

Zhu, T., Xin, F., Wei, S., Liu, Y., Han, Y., et al. (2019). Genomewide identification, phylogeny, and expression profiling of class iii peroxidases gene family in brachypodium distachyon. *Gene* 700, 149–162. doi: 10.1016/j.gene.2019.02.103

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 17 March 2023
DOI 10.3389/fpls.2023.1143512

OPEN ACCESS

EDITED BY
Lida Zhang,
Shanghai Jiao Tong University, China

REVIEWED BY
Liyi Zhang,
Guizhou Academy of Agricultural Sciences (CAAS), China
Christos Noutsos,
State University of New York at Old Westbury, United States

*CORRESPONDENCE
Pradeep Ruperao
r.pradeep@cgiar.org
Abhishek Rathore
abhishek.rathore@cgiar.org

SPECIALTY SECTION
This article was submitted to Plant Bioinformatics, a section of the journal Frontiers in Plant Science

RECEIVED 16 January 2023
ACCEPTED 22 February 2023
PUBLISHED 17 March 2023

CITATION
Ruperao P, Gandham P, Odeny DA, Mayes S, Selvanayagam S, Thirunavukkarasu N, Das RR, Srikanda M, Gandhi H, Habyarimana E, Manyasa E, Nebie B, Deshpande SP and Rathore A (2023) Exploring the sorghum race level diversity utilizing 272 sorghum accessions genomic resources.
Front. Plant Sci. 14:1143512.
doi: 10.3389/fpls.2023.1143512

Exploring the sorghum race level diversity utilizing 272 sorghum accessions genomic resources

Pradeep Ruperao[1*], Prasad Gandham[2], Damaris A. Odeny[1], Sean Mayes[1], Sivasubramani Selvanayagam[3], Nepolean Thirunavukkarasu[4], Roma R. Das[5], Manasa Srikanda[6], Harish Gandhi[7], Ephrem Habyarimana[5], Eric Manyasa[8], Baloua Nebie[9], Santosh P. Deshpande[10] and Abhishek Rathore[11*]

[1]Center of Excellence in Genomics and Systems Biology, International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Hyderabad, India, [2]School of Plant, Environmental and Soil Sciences, Louisiana State University Agricultural Center, LA, United States, [3]Wageningen University and Research, Wageningen, Netherlands, [4]Genomics and Molecular Breeding Lab, Indian Council of Agricultural Research (ICAR) - Indian Institute of Millets Research, Hyderabad, India, [5]International Crops Research Institute for the Semi-Arid Tropics, Hyderabad, India, [6]Department of Statistics, Osmania University, Hyderabad, India, [7]International Maize and Wheat Improvement Center (CIMMYT), Nairobi, Kenya, [8]Sorghum Breeding Program, International Crops Research Institute for the Semi-Arid Tropics, Nairobi, Kenya, [9]International Maize and Wheat Improvement Center (CIMMYT), Dakar, Senegal, [10]Hytech Seed India Private Limited, Hyderabad, India, [11]Excellence in Breeding, International Maize and Wheat Improvement Center (CIMMYT), Hyderabad, India

Due to evolutionary divergence, sorghum race populations exhibit significant genetic and morphological variation. A *k-mer*-based sorghum race sequence comparison identified the conserved *k-mer*s of all 272 accessions from sorghum and the race-specific genetic signatures identified the gene variability in 10,321 genes (PAVs). To understand sorghum race structure, diversity and domestication, a deep learning-based variant calling approach was employed in a set of genotypic data derived from a diverse panel of 272 sorghum accessions. The data resulted in 1.7 million high-quality genome-wide SNPs and identified selective signature (both positive and negative) regions through a genome-wide scan with different (iHS and XP-EHH) statistical methods. We discovered 2,370 genes associated with selection signatures including 179 selective sweep regions distributed over 10 chromosomes. Co-localization of these regions undergoing selective pressure with previously reported QTLs and genes revealed that the signatures of selection could be related to the domestication of important agronomic traits such as biomass and plant height. The developed *k-mer* signatures will be useful in the future to identify the sorghum race and for trait and SNP markers for assisting in plant breeding programs.

KEYWORDS

sorghum race, deep learning, deep variant calling, *k-mer* analysis, selection pressure, gene enrichment, positive and negative selection

Introduction

The process of domestication and natural selection leads to an increased frequency of favorable alleles and subsequently results in complete fixation at target genomic loci (Smýkal et al., 2018). Although the selection process targets advantageous alleles, it also inadvertently results in an increase in the frequency of alleles at neutral loci that are in linkage disequilibrium, a phenomenon referred to as selective sweep (Stephan et al., 1992). A selective sweep has the potential of enhancing the fitness of an individual at the expense of the overall genetic diversity of a population at the respective loci. As a result, modern cultivars are derived from a small fraction of genetically related varieties (Mccouch et al., 2013) in spite of the existence of the vast genetic diversity of global plant germplasm. A better understanding of and stepwise exploitation of existing natural variation in each crop is one key aspect of meeting the increasing food demand in the coming decades.

Sorghum [*Sorghum bicolor* (L.) Moench] is an important cereal crop grown and consumed by a large proportion of the global population. The earliest record of sorghum seeds was recorded at Nabta Playa (Egyptian-Sudanese border) and indicated early domestication (Wendorf et al., 1992). The subsequent migration and adaptation of sorghum across Africa and Asia led to the evolution of morphological and geographically diverse groups, classified into major races (Harlan and Wet, 1972; Harlan and Stemler, 2012). More recent phenotype and genotype-based classifications also support the sorghum race classification within the global diversity panel (Brown et al., 2011). However, inter-racial diversity has not been fully understood in sorghum in a way that allows exploitation of racial structure for heterotic gains. Development of such knowledge would improve overall genomic predictions in sorghum as has been done in other cereal crops (Norman et al., 2018) for the best use of the genome in crop improvement programs.

The extent of genetic diversity is measured by the number of nucleotide variants across individuals and species (Deu et al., 2006; Kebbede, 2020). Such variants range from single nucleotides to large-scale structural differences. However, most studies in the past have only used single nucleotide variation (Afolayan et al., 2019; Enyew et al., 2022) ignoring other structural variations such as insertion-deletions (indels) and presence-absence variations (PAV) (Saxena et al., 2014). PAVs are present in some individuals but absent in others, making them perfect for detecting major differences among multiple genomes. Pangenomes, therefore, can help obtain a more complete set of genomic variants for a species (Hurgobin and Edwards, 2017) since they represent irreversible changes for a given species. The availability of sorghum pangenomes (Ruperao et al., 2021; Tao et al., 2021) makes it possible to carry out a more extensive genetic variation analysis across the different races.

Despite emerging advances in sequencing technologies, distinguishing accurate genetic variants from sequencing errors remains challenging. Because a majority of the genome assembly tools are based on the *de Bruijn* graphs (Zerbino and Birney, 2008; Simpson et al., 2009; Bankevich et al., 2012; Peng et al., 2012), in which the sub-sequence of *k-mer*s (substrings of length *k*) are used to construct the graph and output the paths as contigs (without branching). The resulting contigs can therefore be biased and fragmented as a result of sequencing errors, especially in highly repetitive genomes, leading to low confidence in variant calling. Alternative alignment-free methods of variant detection have been developed using both *k-mer* frequencies and information theory (Song et al., 2014; Pajuste et al., 2017; Zielezinski et al., 2017; Audano et al., 2018). These alignment-free methods have been applied in several studies including for phylogeny estimation (Haubold, 2014), identification of mutations between strains (Nordström et al., 2013) and association mapping (Sheppard et al., 2013).

More recently, deep learning methods have been introduced as a machine learning technique applicable to a range of fields including genomics. Deep learning models can be trained without prior knowledge of genomics and next-generation sequencing (NGS) data to accurately call genetic variants (Telenti et al., 2018). Learning a deep convolutional neural network-based statistical relationship between aligned reads, a genotype calling approach has been implemented in DeepVariant programs (Poplin et al., 2018). The DeepVariant approach is reported to outperform the existing variant calling tools (Poplin et al., 2018).

The objective of our study was to use deep learning (DeepVariant method) to better understand genetic variation, domestication events and selection signatures across known sorghum races. We used existing whole-genome sequence data to quantify genome-wide positive and negative selected regions to enhance our understanding of genome function and the frequency of genetic variations. In addition, we determined the putative signals of selection in sorghum that have resulted from true selective events or population bottlenecks.

Results

DeepVariant calling and annotation

The whole genome sequence (WGS) data (Table 1 and Supplementary Table 1) were mapped (Supplementary Table 2, Supplementary Figure 1) to the sorghum pangenome (Ruperao et al., 2021), and a total of 1.7 million high-quality SNPs, and 470,375 InDels (154,900 insertions, 278,951 deletions and 36,524 mixed variants) were called using the DeepVariant method. Homozygous SNPs were predominant (88.3%) over heterozygous

TABLE 1 A summary of publicly available data used in the current analysis.

Reference	# Genotypes	Average coverage
Valluru et al. (2019)	196	13x
Jensen et al. (2020)	70	7x
Yan et al. (2018)	6	28x
Total	**272**	**16x**

SNPs (11.6%) (Supplementary Table 3). The overall density of SNPs was 2.5 SNP/kb, whereas the indel density was 0.6/kb. The maximum (209,429) and minimum (147,952) SNPs were reported on chromosome 2 (0.3/kbp) and chromosome 9 (0.4/kbp), respectively (Supplementary Table 4) (Figure 1), while the maximum (70,722) and minimum (34,530) number of indels were reported on chromosome 1 (0.8%) and chromosome 8 (0.5%) respectively. Most of the insertions (98%) and deletions (93%) were less than 10bp in length (Supplementary Figure 2A).

SNP annotation reported 11% SNPs of which 51,891 were synonymous and 53,159 non-synonymous, resulting in a non-synonymous-to-synonymous substitution ka/ks ratio of 1.02 (Supplementary Tables 5, 6), consistent with the previous study by Mace et al. (2013). Sorghum accessions NSL54318 (50,238 non-synonymous; 3,933 start gain; 88 start lost and 611 stop gain SNPs) and PI660645 (46520 non-synonymous; 3700 start gained and 103 stop lost) harbored the maximum and minimum effect SNPs (Supplementary Table 5). There were more transitions (C/T and A/G) than transversions (A/T, A/C, T/G and C/G) with a transition/transversion ratio ranging from 1.912 (NSL50716, IS30508) to 1.983 (PI329719). The overall tr/tv ratio was 1.960 (Supplementary Table 7).

SNPs with large effects were the least common (1,362; 0.04%) compared to SNPs with low (63,298; 1.9%), moderate (53,159; 1.6%) and modifying SNPs (96%). A total of 89.3% (1,595,340) of the SNPs were conserved across five sorghum race accessions while the remaining 10.6% (190,321) were variably detected in at least one sorghum race. Among the SNPs in the sorghum race accessions, 0.03% (590) were race-specific, the majority (60.6%; 358) of which were reported in durra and the least in the bicolor race (6.1%; 36) (Supplementary Table 8) (Supplementary Figure 2B). Most of the race-specific SNPs (57.9%) were highly confident with support from more than 10 accessions. Only 21% of the race-specific SNPs were supported by less than 5 accessions (Supplementary Figure 2C).

Sorghum races caudatum, durra, guinea and kafir had the highest proportion of SNPs with the low MAF category (0.0,0.1) compared to bicolor. Kafir had the highest proportion of SNPs with MAF category (0.1, 0.2) while the bicolor race reported the highest proportion of SNPs with MAF greater than 0.2, which is expected for a race with a long history of cultivation (Supplementary Figure 2D).

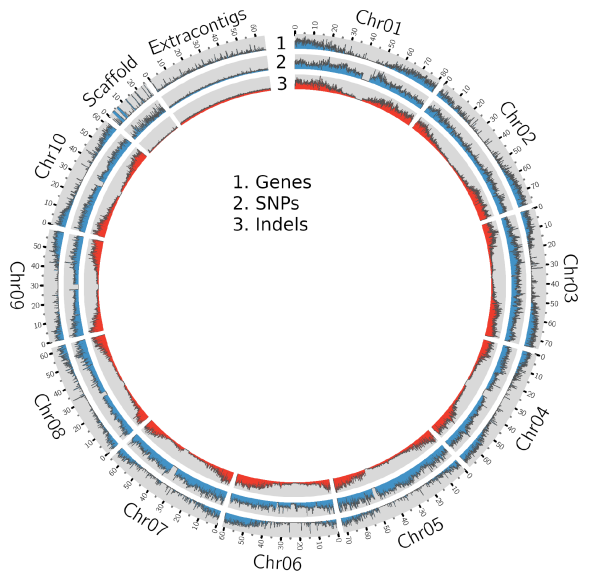


FIGURE 1
A Circos plot showing the density of genes, SNPs and InDels in the sorghum pangenome.

Genetic and nucleotide diversity

The SNP-based Neighbor-Joining (NJ) dendrogram of the 272 genotypes grouped them largely according to race genetic relatedness (Supplementary Figure 3). Four major clusters were observed with a number of subgroups. The phylogenetic tree contained a distinct cluster of 63 guinea race accessions (nodes in blue color) mixed with a few other race individuals, such as durra (PI221662, PI248317, PI267653 and PI148084) (nodes in brown color), kafir (PI660555 and NSL365694) (nodes in pink color), bicolor race (IS12697) (nodes in red color). The other sorghum race clusters were split with non-corresponding sorghum race accessions. For example, durra has 91 accessions split into two clusters with caudatum and kafir accessions. The bicolor accessions were placed mostly in durra and guinea clusters. Among the bicolor accessions, the China origin accessions were grouped distinctly in the durra cluster compared to other bicolor accessions.

The evaluation of nucleotide diversity across all 272 accessions showed that sorghum had low diversity (0.0000483715) compared to wheat (π_A=0.0017, π_B= 0.0025 and π_D= 0.0002) (Zhou et al., 2020), maize (π = 0.014) (Tenaillon et al., 2001) and rice π = 0.0024 (Huang et al., 2010) (Supplementary Figures 4, 5). The diversity varies depending on the population size and the level of diversity of the accessions used in such a population. However, such low diversity was also reported in an earlier study (Sapkota et al., 2020). We observed significant differences ($P < 0.05$) in nucleotide diversity between three sorghum races (caudatum, durra and guinea) that were represented with more than 50 genotypes. The durra had the highest nucleotide diversity while caudatum showed the lowest (π_C= 0.0000419, π_G= 0.0000631 and π_D= 0.0000637). The distribution of nucleotide diversity on the sorghum race genome was in the order of $\pi_D > \pi_G > \pi_C$.

We used the *Fst* index to estimate the temporal genetic divergence between the race accessions and observed that the level of genetic differentiation among the sorghum race populations ranged from moderate (*Fst* = 0.044 for caudatum vs durra) to relatively high (*Fst* = 0.18 for bicolor vs guinea) (Supplementary Tables 9, 10; Supplementary Figure 6) indicating that inter-population differences were relatively low. The average *Fst* between the bicolor and other races was ~0.16, which was higher than in non-bicolor race comparisons suggesting that gene flow from bicolor to other races was much earlier than the gene flow between the rest (non-bicolor) of the races. The durra and guinea

populations revealed the second-highest *Fst* of 0.1228 and were classified as the sorghum race intermediates (Supplementary Table 9). A total of 19,696 SNPs having significant high *Fst* were reported between bicolor-kafir race combinations, of which 910 SNPs were genic SNPs (Supplementary Table 10).

The difference between (diverse) sorghum race populations was measured with Tajima's D (Table 2). A total of 13,070 SNPs were reported to have θ_π (observed value) less than θ_k (expected value) (maximum 4,612 and minimum 1,869 SNPs from durra and bicolor respectively), indicating that the variants may have undergone a recent selective sweep. Another 311,045 SNPs reported greater θ_π compared to θ_k (maximum 202,684 and minimum 76,836 SNPs from guinea and bicolor, respectively) suggesting balancing selection. Compared to non-bicolor race mutations, a lower number of mutations were linked to genes within a selection sweep than with balancing selection genes in the bicolor race (Supplementary Tables 11, 12).

K-mer based divergence

The *k-mer* genetic distance between the sorghum accessions was computed from the size-reduced sketches and distance function developed in the mash tool (Supplementary Table 13). The durra race was the most distinct from the reference pan-genome (Ruperao et al., 2021) based on the mean distance of accessions, followed by guinea (Figure 2A). The bicolor race was the most closely related race to the reference (Figure 2). Accessions from each sorghum race, SCIV4, PI285039, PI276823, PI665088 and PI665108 from bicolor, caudatum, durra, guinea and kafir, respectively were more genetically distinct from the reference (Supplementary Table 13) and representative of the specific race and therefore used for *k-mer* analysis. These distinct sorghum accessions were in agreement with the NJ distance between the accessions (Supplementary Figure 3).

With the optimized 47 *k-mer* size (Figure 2B), the overall *k-mer* sequence comparison between the five race accessions (2.3 billion *k-mers*) showed that 35.3% (434 million unique *k-mers*) of common *k-mers* present in all five races accessions, this indicates the conserved *k-mer* of all sorghum race accessions. The 13.3% (314 million *k-mers*) were commonly seen in any four sorghum race accessions, indicating that these *k-mers* were absent in at least any one of the sorghum races. This variability decreased to 8.8% (108 million *k-mers*) and 6.3% (78 million *k-mers*) on measuring the common *k-mers* between three and two sorghum race accessions respectively. For example, SCIV4 (bicolor) and PI665108 (kafir) shared 402 million distinct *k-mers*, which was 45% and 23.5% of total distinct *k-mers* reported respectively (Figures 2C-D). From this *k-mer* comparison between the sorghum race accessions, 23.8% of *k-mers* were unique to sorghum races. These race-specific *k-mers* were possibly unique to genomic sequence (as a single genome sequence for each race was used for the analysis).

Overall, 10,321 gene PAVs were identified based on the *k-mer* sequence reads mapping to sorghum pan-genome assembly (Supplementary Table 14) (Figure 2E). The mapping of the race-specific *k-mer* sequence reads identified 132, 8009, 211, 445, and 344 unique genes in caudatum, bicolor, guinea, durra, and kafir sorghum accessions, respectively. One hundred and twenty-nine (129) genes were commonly present in all sorghum race accessions (Supplementary Table 15), indicating the *k-mers* are unique with the specific variations or *k-mers* partially mapping the gene length-frequency with horizontal mapping range of 0.4 to 1 (frequency) (Figure 2F). Furthermore, 1,051,453 SNP were identified supporting the *k-mers* sequence (Figure 2G) reads of which, 85,048 SNPs were genic, and 167 SNPs were validated with the SNParray sequences (Figure 2H) (Supplementary Table 16) used for sorghum pangenome analysis (Ruperao et al., 2021).

Selection signatures

Several sweep regions were detected with iHS (Figures 3A, B and Supplementary Figure 3), of which, 64 were significant (FDR < 0.05) (Supplementary Table 17). The majority of sweeps were reported on chromosome 7 (19 regions) followed by chromosome 4 (17 regions) and chromosome 10 (2 regions) (Supplementary Table 17). The highest number of selective sweep regions were observed in durra (54 regions), followed by caudatum (51), guinea (45), kafir (38) and bicolor (30) (Supplementary Tables 18, 19). A total of 14 selective sweep regions were common in all five sorghum races while 21 regions were uniquely absent in any one sorghum race (Supplementary Table 18). For example, 9 selective sweep regions were reported in four sorghum races but uniquely absent in the bicolor race alone (Supplementary Table 18).

We used the cross-population extended haplotype homozygosity (XP-EHH) score and detected sweep regions from each combination of sorghum race population (Supplementary Figure 7) (Table 3). We identified 8,888 significant (FDR < 0.05) selection sweep regions, of which 3,504 regions were common between more than

TABLE 2 Summary SNP statistics in Sorghum race populations.

Race	SNP	π (10^{-5})	Tajima's D ($\theta\pi > \theta k$)	Non-synonymous SNPs	Synonymous SNPs	Non-synonymous/ Synonymous
Bicolor	374,545	3.37	0.44	11,100	9,957	1.114
Caudatum	961,644	2.72	0.27	31,391	31,168	1.007
Guinea	976,558	4.30	0.37	31,827	31,681	1.004
Durra	1,261,078	3.89	0.18	38,362	37,651	1.018
Kafir	886,880	4.72	0.26	27,779	27,458	1.011

10.3389/fpls.2023.1143512

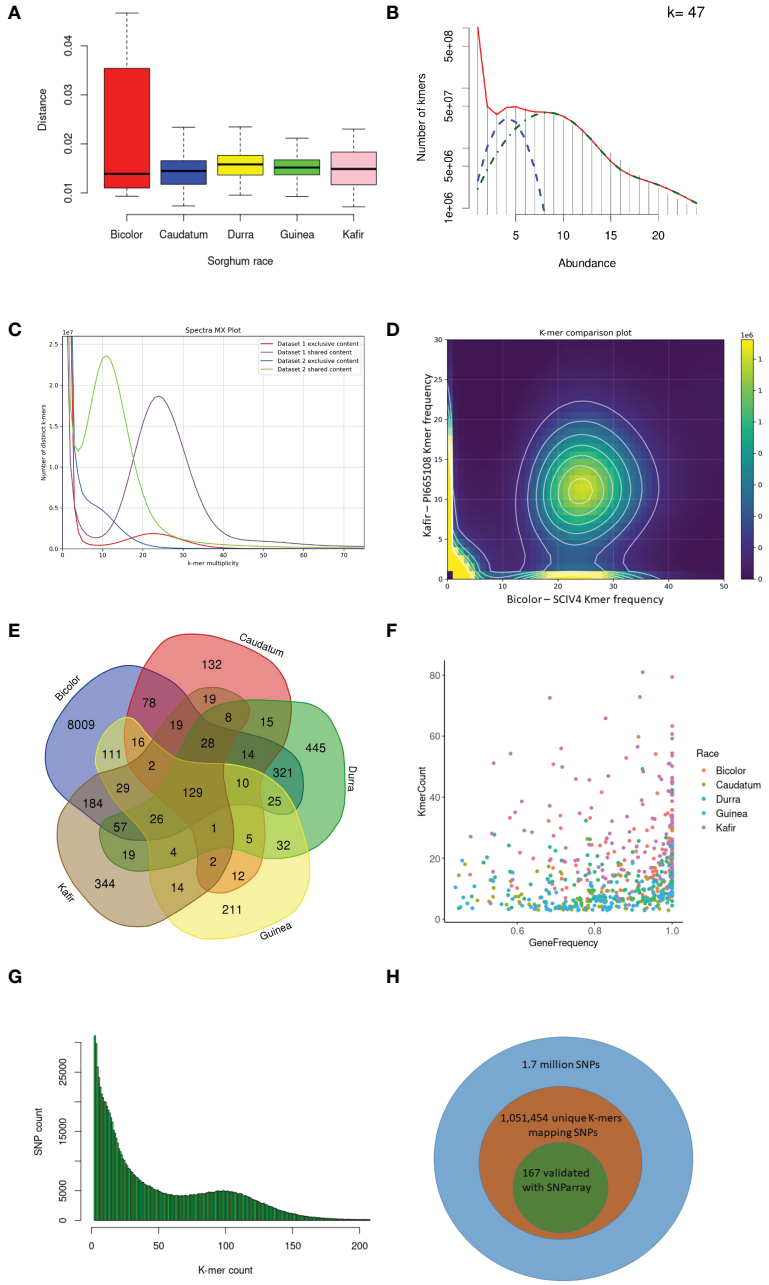


FIGURE 2
K-mer and read mapping overview. **(A)** An alignment-free method, the Jaccard index uses the hash procedure to measure the distance between sorghum race accessions. **(B)** Sampled histogram and fit for 47 *k-mer* lengths. Red is the fit of the complete statistical model of the histogram, blue is the heterozygous *k-mers* and green is the only homozygous *k-mers* **(C)** The *k-mers* share between kafir (PI665108) and bicolor (SCIV4) sorghum race accessions as dataset1 and dataset2 respectively. **(D)** Distinct *k-mer* share between kafir and sorghum race, the cloud indicates the shared *k-mers* and heigh density *k-mers* on x and y-axis are unique *k-mers* respectively. **(E)** Mapped *k-mer* sequence reads in the number of genes in each sorghum race accessions **(F)** Proportion of sorghum race common genes covered with horizontal and vertical coverage with *k-mer* sequence reads. **(G)** The unique *k-mers* holding the deepvariant SNPs with respective *k-mer* count and off these, **(H)** the proportion of deepvariant SNPs validated.

two sorghum race combinations (Supplementary Tables 20-S29). Out of all selective sweeps identified from the sorghum race combinations, chromosome 5 had the maximum of 1,399 regions while chromosome 9 had the least (616). The kafir population exhibited the highest (2,473) selective sweep regions in comparison with the guinea race (Supplementary Table 28). Only a few (525) sweep regions were reported in bicolor population (Supplementary Table 23).

Overlapping selection regions between Tajima's D and XP-EHH

We defined the overlapping selection regions as those located beyond the thresholds and in the same chromosome sequence location. Tajima's D statistics were obtained from each sorghum race population dataset and identified the genes which did not fit the neutral theory model at equilibrium between mutation and

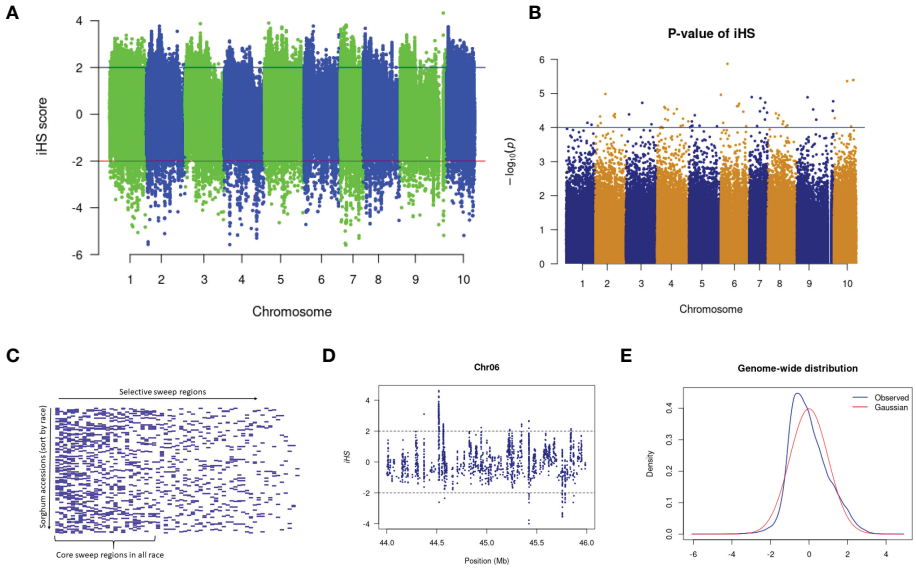


FIGURE 3
(A) The iHS score distribution across the genome with **(B)** associated p-values. **(C)** Selective sweep regions in sorghum accessions sorted by race. **(D)** A subset of iHS on chromosome 6 showing the regions having values above and below the threshold value and **(E)** The distributions of the standardized iHS scores and comparison with standard Gaussian distribution.

TABLE 3 Description of the candidate selective sweep regions detected using XP-EHH between the sorghum race populations.

Sorghum race combination	Significant XP-EHH regions	Supporting iHS	XP-EHH candidate genes$	XP-EHH mapping genes&
Bicolor × Caudatum	671	3	212	109
Bicolor × Durra	640	3	183	127
Bicolor × Guinea	612	5	201	127
Bicolor × Kafir	522	0	117	61
Durra × Caudatum	1,832	1	587	414
Durra × Guinea	2,350	4	677	367
Durra × Kafir	1,761	3	424	356
Guinea × Caudatum	2,084	6	607	328
Guinea × Kafir	2,070	1	711	426
Kafir × Caudatum	1,263	2	425	365

$ The number of genes refers to the genes mapping 5kb upstream/downstream to selective sweep region. & The selective sweep region present within the gene regions.

genetic drift. A total of 324,115 genome-wide bins were observed with non-equilibrium statistics of neutrality test, of which 311,045 (with SNPs in range of 76,836-bicolor to 202,684-guinea) were undergoing purifying selection (negative selection) and 13,070 were (with SNPs in range of 1,869-bicolor to 4,612 durra) selection maintained (balanced positive selection) (Figure 4A). Among the variants undergoing purifying selection, 43,191 bins had a significant low *Fst* index supporting the signature of a recent population expansion (Figure 4B), of which 14% were from genic regions (Supplementary Table 30). The purifying selection regions had low diversity (Figure 4D) with reduced allele frequency in the descendant population compared to the ancestral population.

The significant selection regions (FDR <0.05) detected by XP-EHH were specific to the pair-wise sorghum race combinations. Among the identified 8,888 significant XP-EHH candidate-sweep regions from overall sorghum race combinations, of which 179 regions were genic (Supplementary Table 31) and "selection-maintained variations" indicating the recent population contraction. The overall sweep regions were in the comparable range identified in other crops such as wheat, (3,105 – 16,141 sweep regions in the genome of domesticated einkorn and emmer lines; Zhou et al., 2020) and soybean (3,811 genes positioned within the selective sweep regions)(Saleem et al., 2021). Among the chromosomes, chromosome 5 and groups of 4,7,9,10

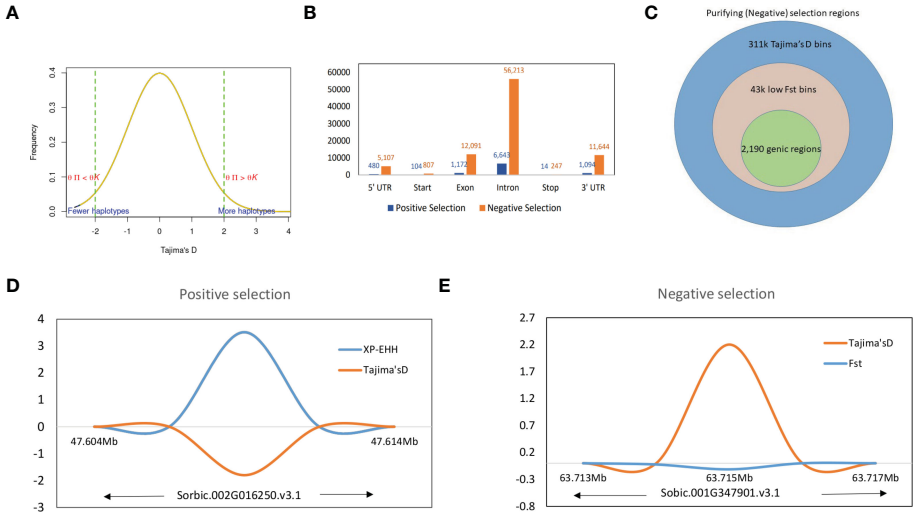


FIGURE 4
(A) Tajima's D values distribution with signifying positive and negative thresholds. **(B)** Structural annotations of genomic regions in positive and negative thresholds. **(C)** The proportion of higher Tajima's D values and lower *Fst* valued bins as negative selection regions and corresponding genic regions. **(D)** The positive selection region on Chr09 with XP-EHH score and Tajima's D valued plot and **(E)** A genic negative selection region on Chr01 (specific to guinea race) with significantly higher Tajima's D and lower *Fst* value regions.

chromosomes contained the highest (33) and lowest (14) numbers of genes, respectively. A relatively low Tajima's D was observed in selective sweep regions when compared with a significantly higher XP-EHH valued region (Figure 4D).

Enrichment of candidate genes under selection

A total of 2,370 genes genome-wide were observed to deviate significantly using equilibrium/neutrality tests, of which 179 were selection-maintained (balanced selection) while 2,191 were undergoing purifying selection (Supplementary Tables 30, 31). Durra and bicolor had the maximum (110) and minimum (39) number of genes undergoing positive selection respectively. Bicolor (409) and guinea (1,133) had the maximum and a minimum number of selection-maintained genes.

A similar trend of the fewest number of genes were reported in the bicolor race (421), with guinea having the maximum (1,166) genes under purifying selection. Among the five races, guinea and kafir shares the maximum number of common selection-maintained (26) and purifying (70) genes, suggesting potential rich gene flow between these two races (Figure 5). Additionally, guinea, kafir and durra reported maximum genes as sweep regions

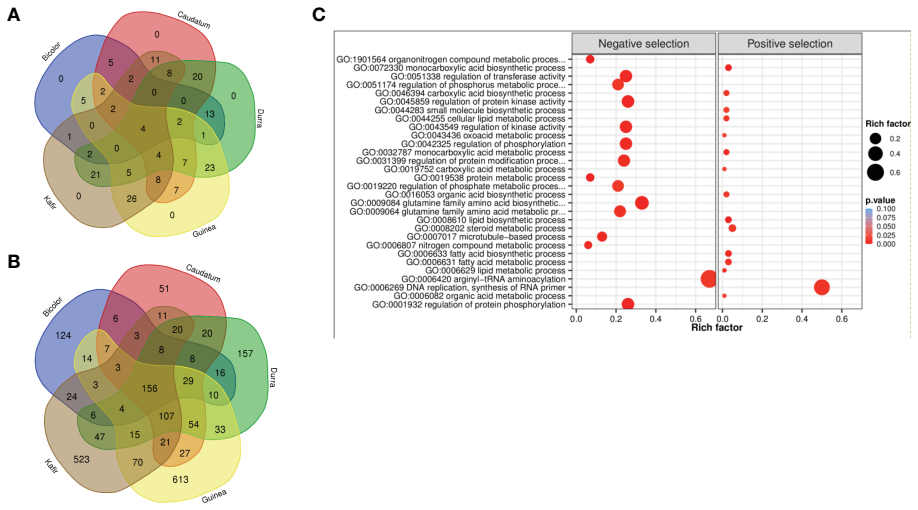


FIGURE 5
Venn diagram of **(A)** positive selection and **(B)** negative selection genes from five sorghum races combinations. **(C)** Significantly enriched GO biological terms for positive and negative selected genes (bubble color indicates the p-value range and size indicates the gene ratio).

(guinea-kafir: 26, durra-guinea: 23 and durra-kafir: 21) (Figure 5A), with low nucleotide diversity (π) in caudatum and bicolor (Figure 3B) also indicating the traits regulated by these regions may have undergone similar histories of selection.

The 2,370 genes undergoing selection pressure (both positive and negative) showed significantly enriched gene ontology (GO) term and among these genes (Figure 5C, Supplementary Figure 8), the top GO term was lipid biosynthetic process (GO:0008610) and organonitrogen compound metabolic process (GO:1901564) for genes with positive and negatively selection, respectively (Supplementary Table 32). Among the positively selected gene set, most of them were enriched with lipid biosynthetic process (GO:0008610), metabolic process (GO:0006629), carboxylic acid metabolic process (GO:0019752), oxoacid metabolic process (GO:0043436), organic acid metabolic process (GO:0006082), most of these metabolic pathways were related to plant stress resistance. Whereas the negatively selected gene was majorly enriched with nitrogen compound metabolic process (GO:0006807), organonitrogen compound metabolic process (GO:1901564) and protein metabolic process (GO:0019538) (Supplementary Table 32). The nitrogen utilization and metabolic pathway were found significantly enriched and confirmed the genes under selection throughout either domestication or during subsequent breeding with earlier selection study (Massel et al., 2016). The genes enriched with 'DNA replication, 'lipid metabolism' and 'hormone signal' suggest that sorghum has evolved defense strategies, and enrichment of phosphorylation, kinase activity, transferase, phosphate and phosphorus metabolic process triggers many metabolic processes and plant growth activity.

Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were identified according to the selection's signature candidate gene with a p-value <0.05 (Figure 6). A KEGG pathway enrichment analysis was performed for the selection signature gene to identify the number of significantly changed samples along the pathway that were relevant to the background number. A total of 2,370 genes were mapped onto 315 pathways, and the most enriched sequences were metabolic pathways and biosynthesis pathways. The top 14

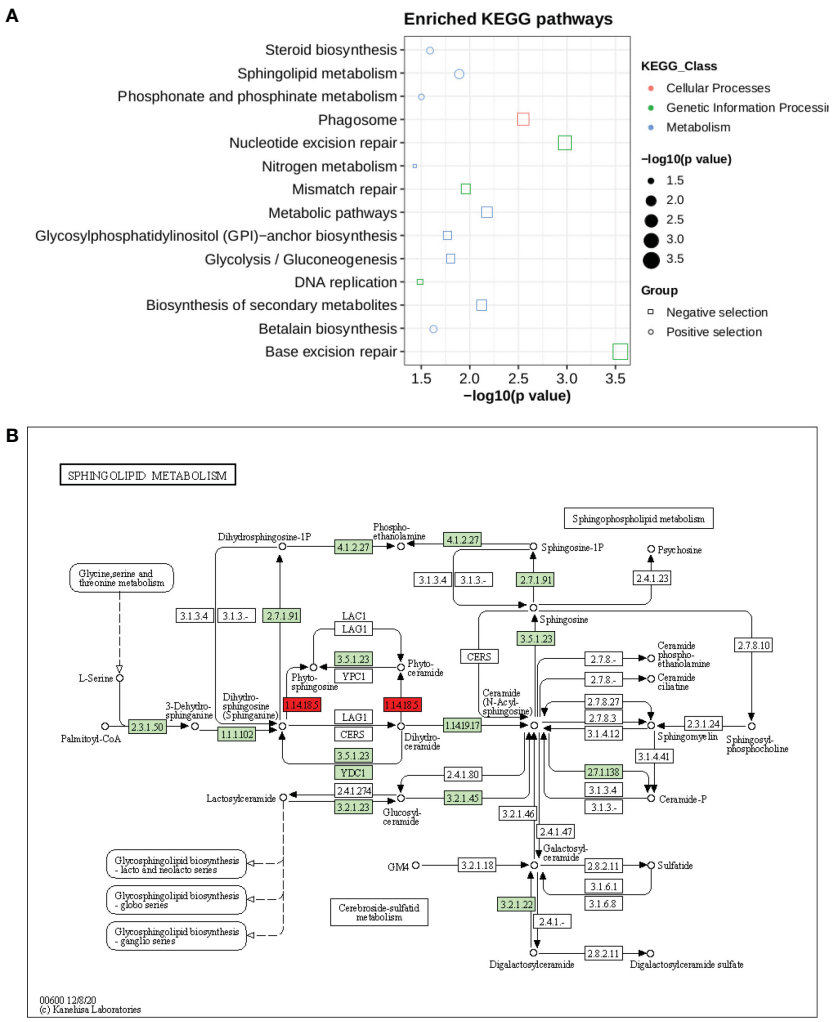


FIGURE 6
(A) KEGG pathway enrichment for genes under selection pressure **(B)** Pathway of sphingolipid metabolism. The two red boxes represent the positively selected genes involved in pathways.

pathways with the greatest number of annotated sequences are shown in Supplementary Table 33. Most of the significant pathways were in metabolism, biosynthesis, excision repair and secondary metabolites. The most significantly changed KEGG pathways were in sphingolipid metabolism (Figure 6B), betalain, steroid biosynthesis, phosphonate and phosphinate pathways for positive selection genes. Sphingolipids are essential components of plasma membrane providing structural integrity to plant membrane, regulating the cellular process, and also enhancing the tolerance of sorghum to biotic and abiotic stresses. Steroid hormone biosynthesis and the phosphonate and phosphinate metabolism pathways are also involved in the adaptation of sorghum to low salinity. Whereas base and nucleotide excision repair, biosynthesis of secondary metabolites, glycolysis, GPI and nitrogen metabolism were significantly enriched in negative selection genes (Supplementary Table 33). These annotations provide valuable information for studying the specific biological and metabolic processes and functions of genes under selection pressure in sorghum accessions.

Overlap of signatures of selection with QTLs

Quantitative trait loci associated with seven traits that overlapped with detected signatures of selection were compared with earlier reported sorghum QTLs (Hostetler et al., 2021). Analysis of the overlaps between signatures of selection and reported QTL indicated that 10 and 206 linked genes were identified as positively and negatively selected genes respectively (Supplementary Tables 34, 35). Some QTL for traits of plant height, root biomass, dead above-ground biomass, live above-ground biomass and total biomass overlapped significantly with putative gene regions of signatures of selection.

Discussion

We have demonstrated the utility of vast the sorghum genomic data that exists in public databases for characterization of a representative set of sorghum (Valluru et al. 2019). Our results validate the application of deep learning for the characterization of sorghum races and goes further to establish nucleotide diversity and genetic divergence across and within different sorghum races. We also used existing QTL data to identify candidate genes that are under both negative and positive selection.

The sorghum reference set used in the current study was earlier selected by Billot et al. (2013) after genotyping 3,367 global collections using 41 representative nuclear SSR markers and is considered to be representative of the sorghum collections that exist in various global gene banks. Our results confirm that the clustering of sorghum germplasm was largely according to regions (Paterson et al., 2009; Bekele et al., 2013; Ramu et al., 2013) indicating continuous gene flow between various racial groups depending on where the sorghum races are grown. The only distinct exception was in the guinea race, which was expected since the guinea race is specifically grown in West Africa and therefore any gene flow would be confined within the West African locations.

Our study also identified many intermediate accessions (more than 15 accessions) as a result of the continuous gene flow suggesting that a different criterion other than morphology will need to be used in future studies for the correct classification of sorghum races and their intermediates.

Capturing race-specific sequences will be critical in future studies for the follow-up identification of variants and/or, genes associated with each sorghum race. For example, longer k-mers (>15 bp) have been utilized as biomarkers (Drouin et al., 2016; Wang et al., 2018) as they can hold biological information and depict specific signatures in nucleotide sequences (Wang et al., 2016). Our ability to differentiate abundant k-mers between the different sorghum races in the current study provides an opportunity for future studies to utilize k-mers as race- or accession-specific identifiers in sorghum. Currently we were able to identify the sorghum race-specific k-mers, that are present in respective race, and able to locate the position and associate the genomic features. Based on the sequence read mapping, the gene PAV was earlier identified in the sorghum pangenome (Ruperao et al., 2021), and adapting the similar approach the genes having the race specific k-mers regions were also reported. With the known unique k-mer position, it is possible to extend the study of the genomic features having race-specific unique sequence (such as any genetic variations including the SSR, SNP, CNV and SV).

DeepVariant calling and its utility in sorghum breeding

For the first time in sorghum, we used the DeepVariant (Poplin et al., 2018) tool, a deep learning approach for SNP calling, and reported over two million genome-wide variants from existing sequencing data. One of the concerns of SNP calling from NGS data is the accuracy of SNPs. A recent comparison of SNPs called from the traditional SNP calling tools such as GATK (Depristo et al., 2011) with DeepVariant method reported superior performance of the latter (Lin et al., 2022), further validating our choice to implement this method in sorghum. Our results were largely consistent with previous studies in sorghum that involved SNP calling from NGS data, including the patterns of SNP distribution observed across the genome (Paterson et al., 2009; Bekele et al., 2013) and the non-synonymous to synonymous SNP substitution ratio. In this study, the deepvariant has called the variants with 0.19/Kbp which is comparatively less dense than the earlier reported results (0.33/Kbp) with GATK (Ruperao et al., 2021). Our results were also within the range reported for other genome-wide studies such as in soybean (Lam et al., 2010), rice 1.2 (McNally et al., 2009), and Arabidopsis 0.8 (Clark et al., 2007).

Future studies will need to compare DeepVariant with other existing methods and validate our results in different germplasm sets, such as the sorghum diversity panel (Casa et al., 2008). Such future studies will also need to pay special attention to sequence coverage and how it would affect the accuracy of variants called.

Our study used a minimum overall coverage of 5x, which was more than adequate even for a less efficient SNP calling pipeline (Wu et al., 2019). Sequence coverage is one of the major factors affecting the accuracy of SNPs called from NGS datasets, especially in heterozygous species (Gong and Han, 2022). A coverage of 0.01x has been reported as the most cost-effective coverage in sorghum, with 94.1% SNP accuracy (Jensen et al., 2020). There will be a need for additional studies establishing the effect of various levels of coverage in the NGS datasets for DeepVariant calling, and how it would affect the SNP accuracy in sorghum.

Nucleotide diversity and divergence in sorghum

The genetic relatedness from the NJ tree (Supplementary Figure 3) and Pco (Supplementary Figure 9) analysis between the sorghum race accession demonstrates most of the guinea accessions forms the cluster, except for few accessions relates to caudatum race. Whereas durra race represented in two clusters, one cluster close to guinea and second in between kafir and caudatum (Supplementary Figure 3). Such intermediate race accessions and split of sorghum race clustering was also seen earlier for 389 sorghum diverse panel (Sapkota et al., 2020). On further investigation of structure analysis supports the two subpopulation clusters (K=2) in the sorghum population (Supplementary Figure 10) supporting the distance-based NJ analysis, indicating that the race accessions are genetically related. Our study was purely based on existing data and did not allow for much flexibility in the number of genotypes per race. The overall nucleotide diversity observed for sorghum of π = 0.000048 is significantly smaller than previously reported by Faye et al. (2019) but comparable to a more recent study (Sapkota et al., 2020) that reported π = 0.000032. This figure is much lower than for other cereals such as wheat (π_A=0.0017, π_B = 0.0025 and π_D = 0.0002) (Zhou et al., 2020), maize (π = 0.014) (Tenaillon et al., 2001) and rice (π = 0.0024; Huang et al., 2010) and could be a consequence of the limited number of genotypes used in the study. The race-specific nucleotide diversity indicated that the caudatum (57 genotypes; π_C = 0.0000419) had the lowest diversity followed by guinea (68 genotypes; 0.0000631) and durra (82 genotypes; π_D = 0.0000637) races. On comparing the linkage disequilibrium (LD) decay, rapid LD decay was observed in durra followed by guinea and caudatum (Supplementary Figure 10), supporting the above diversity values of the sorghum race. The least diversity race population (caudatum) shows the higher extents of LD than the races with higher diversity (Durra). Supporting to these results, caudatum race consistently demonstrate the least genetically diverse showed higher LD values (Sapkota et al., 2020). However, studies also reported, the guinea race as the most genetically diverse sorghum type (Morris et al., 2013; Faye et al., 2019). Comparing our results and those of Morris et al. (2013) and Faye et al. (2019), suggests a positive correlation between the number of genotypes per race with the nucleotide diversity. More studies need to be done to confirm the effective population size per sorghum race that will be optimum for a reliable and consistent nucleotide diversity result.

Selection signatures

We used two approaches to detect selection sweeps across the sorghum genome, both of which are haplotype-based. The iHS method, which is based on a single population, was meant to detect recent positive selection (Voight et al., 2006), while the XP-EHH is based on the comparison of two populations and is considered powerful in detecting beneficial alleles shortly before, or at fixation (Alexandra et al., 2015). Such multiple statistical approaches were earlier used for selection sweeps in other crops like cotton (*Gossypium herbaceum*) (Nazir et al., 2020) and soybean (*Glycine max*) (Zhong et al., 2022). A recent study comparing different methods used for detecting selection sweeps reported that both iHS and XP-EHH were able to identify genomic regions undergoing selective sweep under a wide range of population structure scenarios (Vatsiou et al., 2016). Previous studies in sorghum have also reported evidence of selective sweeps in sorghum (Casa et al., 2006; Faye et al., 2019) although the methods used for detection were different. Our results on selective sweep regions were further strengthened by Tajima's D results, which enabled us to identify candidate genes in the significant selective sweep regions.

The 2,370 candidate genes identified in our study (for under selection pressure), of which, 7.5% are positively selected, are similar to the proportion of genes identified for domestication and improvement using the gene-based population study by Mace et al. (2013). The genomic regions that are either positively or negatively selected in the respective sorghum races could give a hint on geographic preferences. More studies will need to delve deeper into specific regional selection sweeps that could eventually be used to predict ideal genotypes/phenotypes. The remaining candidate genes that were reported as undergoing negative selection with evidence from both Tajima's D and Fst index values. Such genomic analysis of crop landraces would enhance our understanding of the basis of local adaptions (Li et al., 2017; Swarts et al., 2017).

Some of the trait-associated genes undergoing selection pressure that have been reported include the dry pithy stem gene mutation that led to the origin of sweet sorghum (Zhang et al., 2018), local adaptation to parasite pressure and signatures of balancing selection surrounding low germination stimulant (Bellis et al., 2020) and the strong selection pressure on the sorghum maturity gene (Ma3) (Wang et al., 2015). Comparative population genomics assist in dissecting the domestication and genome-wide effects of selection as studied in cotton, with reports that 311 selection sweep regions are associated with domestication and improvement (Nazir et al., 2020) and with selection sweeps identified comparing wild and domesticated soybean accessions (Zhong et al., 2022).

Populations subjected to strong selection pressure may experience genetic bottlenecks and result in a loss of genetic diversity. The level of diversity preserved in a population depends on the background of the emerging adaptive alleles (Wilson et al., 2017). Identification of such a large number of selection sweeps suggests the existence of domestication bottlenecks. The identified selection sweeps overlapped with highly differentiated regions suggesting the occurrence of differentiation due to human-mediated selection. These regions help in understanding the

genetic basis of domestication and improvement in traits. On further comparison of the selection regions with significant loci of GWAS analysis (narrowing down the region), it may be possible to determine the genes underlying domestication and selection in the sorghum crop.

The results from this study lead to a better understanding of the changes at the genomic level caused by domestication, selection and improvement of sorghum accessions.

Methods

Plant material

We used 272 sorghum accessions, which included accessions that had been used in a previous sorghum pangenome study (Ruperao et al., 2021) and six sorghum bicolor accessions reported in Yan et al. (2018) (Supplementary Table 1). Among these genotypes, 82, 21, 57, 68, 14 were durra, kafir, caudatum, guinea and bicolor respectively, while the remaining were mixed accessions.

Variant discovery

The fastq sequence reads generated from the 272 sorghum accessions were trimmed with Trimmomatic 0.39 (Bolger et al., 2014). Alignments to the sorghum pangenome (dataverse.icrisat.org, https://doi.org/10.21421/D2/RIO2QM) as a reference (Ruperao et al., 2021) were performed using Bowtie2 version 2.4.2 (Langmead and Salzberg, 2012). All alignments were converted to binary files with Samtools 1.13 (Li et al., 2009) followed by filtering out the read duplication with Picard tools (http://broadinstitute.github.io/picard). The open-source DeepVariant (https://github.com/google/&6;deepvariant) (Poplin et al., 2018) tool was used to create individual genome call sets, followed by merging call sets with Bcftools 1.9 (Bcftools by samtools) then analyzing the merged call set. The merged variants were filtered with 'maf 0.01 min-meanDP 2 minQ 20'. Filtering was done using Vcftools 0.1.16 (Danecek et al., 2011). Retained high-quality sites were used for downstream analysis. Functional annotation of SNPs was done using SnpEff v.4.3 (Cingolani et al., 2012).

Counting *k-mers*

The *k-mer-based* genetic distance between 272 sorghum accessions was measured with Mash (Ondov et al., 2016). Out of the 272 accessions used, the mean distance values within each race were used to compare *k-mers* between the sorghum races. To compare sequences across sorghum races, we determined *k-mer* frequency in sequencing reads from all samples. To identify the common and unique genomic sequences between the sorghum races, we split the sequencing reads into *k* length of the sequence. The optimal *k-mer* size for identifying the distinct *k-mers* was estimated using KmerGenie (Chikhi and Medvedev, 2014) within the *k* range of 21 to 121. The optimized *k*=47 was used for measuring the *k-mer* frequency as shown in the Figure 7. We used the hash-based tool Jellyfish (Marçais and Kingsford, 2011) to count *k-mers* with the optimized *k-mer* length of 31 (kmer 31, expect number of k-mers 100G, count both strand canonical representation, number of threads 25, number of files open simultaneously 2, output file name) and filtered out *k-mers* that appeared only once in samples as they were likely from sequencing errors. The *k-mer* hashes were visually inspected through KAT density plots (Mapleson et al., 2017) for all five sorghum race accessions by producing the *k-mer* frequency, GC plots and contamination checks. Unique *k-mers* mapped to sorghum pangenome were validated with mapped SNParray region from a previous study (Ruperao et al., 2021). Based on the Bowtie2 v2.4.2 (Langmead and Salzberg, 2012) mapped *k-mers*, the gene coverage was assessed with samtools mpileup (Li et al., 2009). The sequence region supporting with minimum of three k-mers was considered as sequence region present in the genome. The gene PAVs were extracted from the sorghum pangenome genes PAVs catalog (Ruperao et al., 2021) with in-house developed script.

Nucleotide diversity and relatedness

The filtered SNPs were further subgrouped based on race-specific variant alleles. Nucleotide diversity (*pi*) was calculated using Vcftools 0.1.16 (Danecek et al., 2011). The *pi* (π) distributions were compared to assess changes in genetic diversity over time. The *pi* (π) density plots were generated with in-house developed scripts.

In addition, a 1,000 bootstrap resampling was used to estimate the genetic relationship among the accessions with R "ape" (Paradis et al., 2004) package to construct a NJ tree and visualized it in iTOL tree viewer (Letunic and Bork, 2019). The Pco analysis was done with R "labdsv" package (https://CRAN.R-project.org/package=labdsv). The admixture v1.3.0 (Alexander et al., 2009) was used to estimate the population structure enabling the cross-validation (CV) with –cv flag. The cross-validation procedure was performed to 10-fold and the lowest CV was considered as optimal K value and the results were visualized with R package (github.com/royfrancis/pophelperShiny) POPHELPER v2.1.1 (Francis, 2017). PopLDdecay (Zhang et al., 2019) was used with MAF 0.01 and MaxDist 2000 to generate the linkage disequilibrium stats and Plot_MultiPop.pl used for plotting the LDdecay.

Population differentiation and signatures of selection

Tajima's D (Tajima, 1989) and per-site *Fst* (based on Weir and Cockerham's *Fst* estimator) (Weir and Cockerham, 1984) were calculated using Vcftools 0.1.16 software (Danecek et al., 2011). Integrated haplotype score (iHS) (Voight et al., 2006) analysis was performed using the "rehh" package (Gautier et al., 2017) in R v 3.6.3, while the extended haplotype-based homozygosity score test)

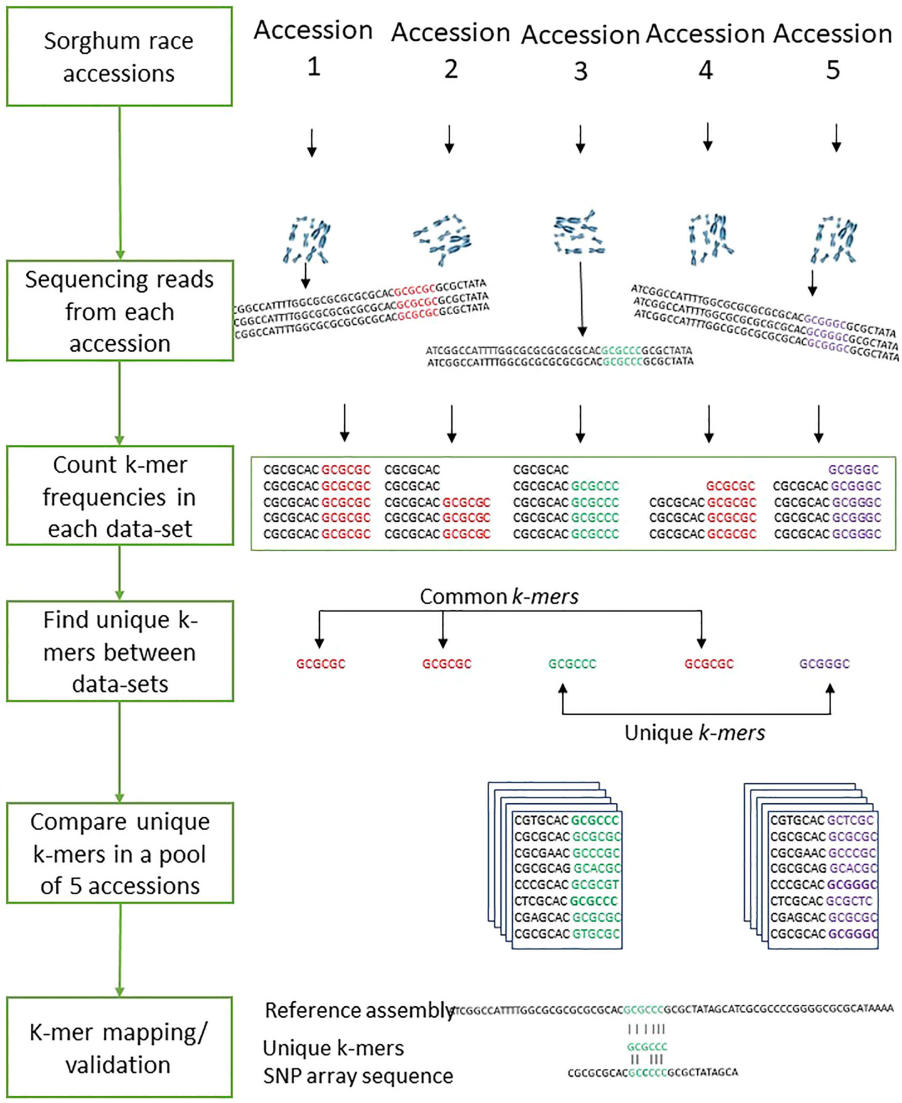


FIGURE 7
The workflows of the sorghum race accessions comparison with *k-mer* analysis.

(XP-EHH) (Sabeti et al., 2007) was derived using Beagle (Browning et al., 2021). Significant selective sweeps were detected using the Bonferroni FDR threshold ($P < 0.05$).

Overlap of putative genomic regions under selection with previously known QTLs was detected after downloading the mapped QTL regions from Hostetler et al. (2021) and comparing them with the identified selection regions.

Access to raw data

We obtained publicly accessible raw Illumina sequence data from three previous studies as shown in Table 1 and Supplementary Table 1. The sorghum accessions having minimum 5x coverage of whole-genome sequence data were used for the analysis, resulting in a total of 272 sorghum accessions.

Conclusion

This study compared the genomes of the sorghum races with short *k-mer* length sequence to identify the conserved and signature patterns of sorghum race sequences. We implemented a deep learning method to detect the variants and compared structural and functional annotations. On applying the *k-mer*-based genome comparison among the sorghum races, we were able to identify the unique *k-mer* sequences that is specific to the sorghum races and also possibly use as race-specific or accession specific (if *k-mer*s compared between accessions) genetic markers. Our study observed a relatively lower genetic diversity in the caudatum and bicolor races than in kafir, guinea and durra races. Our results revealed several putative footprints of selection that harbor interesting candidate genes associated with agronomically important traits using different statistical approaches. The findings will enhance

our understanding of the dynamics of the sorghum race genomes and help to design strategies to breed better genotypes.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

Author contributions

PR conceived the research. PR, PG, MS, and RD carried out the research. SS managed the computational resource. SD and DO provided inputs on sorghum history and domestication. NT, EH, EM, SM, and BN assessed and provided inputs for manuscript. PR, NT, AR, and DO drafted and edited the manuscript. All authors contributed to the article and approved the submitted version.

Funding

The authors also acknowledge the supporting funds from AVISA (OPP1198373) and ICAR-BMGF (101165). We also acknowledge the support from the Bill and Melinda Gates Foundation (BMGF – INV-037010).

Conflict of interest

SPD joined Hytech Seed India Private Limited at the time of manuscript preparation.

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1143512/full#supplementary-material

References

Afolayan, G., Deshpande, S. P., Aladele, S. E., Kolawole, A. O., Angarawai, I., Nwosu, D. J., et al. (2019). Genetic diversity assessment of sorghum (Sorghum bicolor (L.) moench) accessions using single nucleotide polymorphism markers. *Plant Genetic Resources*. Cambridge: Cambridge University Press. 17 (5), 412–420. doi: 10.1017/S1479262119000212

Alexander, D. H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19 (9). doi: 10.1101/gr.094052.109

Audano, P. A., Ravishankar, S., and Vannberg, F. O. (2018). Mapping-free variant calling using haplotype reconstruction from k-mer frequencies. *Bioinformatics* 34 (10). doi: 10.1093/bioinformatics/btx753

Bcftools by samtools.

The genetic organization of sorghum.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021

Bekele, W. A., Wieckhorst, S., Friedt, W., and Snowdon, R. J. (2013). High-throughput genomics in sorghum: from whole-genome resequencing to a SNP screening array. *Plant Biotechnol. J.* 11, 1112–1125. doi: 10.1111/pbi.12106

Bellis, E. S., Kelly, E. A., Lorts, C. M., Gao, H., DeLeo, V. L., Rouhan, G., et al. (2020). Genomics of sorghum local adaptation to a parasitic plant. *Proc. Natl. Acad. Sci. U. S. A.* 117 (8). doi: 10.1073/pnas.1908707117

Billot, C., Ramu, P., Bouchet, S., Chantereau, J., Deu, M., Gardes, L., et al. (2013). Massive sorghum collection genotyped with SSR markers to enhance use of global genetic resources. *PloS One* 8 (4). doi: 10.1371/journal.pone.0059714

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Brown, P. J., Myles, S., and Kresovich, S. (2011). Genetic support for phenotype-based racial classification in sorghum. *Crop Sci.* 51, 224–230. doi: 10.2135/cropsci2010.03.0179

Browning, B. L., Tian, X., Zhou, Y., and Browning, S. R. (2021). Fast two-stage phasing of large-scale sequence data. *Am. J. Hum. Genet.* 108, 1880–1890. doi: 10.1016/j.ajhg.2021.08.005

Casa, A. M., Mitchell, S. E., Jensen, J. D., Hamblin, M. T., Paterson, A. H., Aquadro, C. F., et al. (2006). Evidence for a selective sweep on chromosome 1 of cultivated sorghum. *Crop Sci.* 46 (SUPPL. 1). doi: 10.2135/cropsci2006.0001tpg

Casa, A. M., Pressoir, G., Brown, P. J., Mitchell, S. E., Rooney, W. L., Tuinstra, M. R., et al. (2008). Community resources and strategies for association mapping in sorghum. *Crop Sci.* 48 (1). doi: 10.2135/cropsci2007.02.0080

Chikhi, R., and Medvedev, P. (2014). Informed and automated k-mer size selection for genome assembly. *Bioinformatics* 30, 31–37. doi: 10.1093/bioinformatics/btt310

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

Clark, R. M., Schweikert, G., Toomajian, C., Ossowski, S., Zeller, G., Shinn, P., et al. (2007). Common sequence polymorphisms shaping genetic diversity in arabidopsis thaliana. *Sci. (1979)* 317, 338–342. doi: 10.1126/science.1138632

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., et al. (2011). The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158. doi: 10.1093/bioinformatics/btr330

Depristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* 43 (5). doi: 10.1038/ng.806

Deu, M., Rattunde, F., and Chantereau, J. (2006). A global view of genetic diversity in cultivated sorghums using a core collection. *Genome* 49(2), 168–80. doi: 10.1139/g05-092

Drouin, A., Giguère, S., Déraspe, M., Marchand, M., Tyers, M., Loo, V. G., et al. (2016). Predictive computational phenotyping and biomarker discovery using reference-free genome comparisons. *BMC Genomics* 17, 1–15. doi: 10.1186/s12864-016-2889-6

Enyew, M., Feyissa, T., Carlsson, A. S., Tesfaye, K., Hammenhag, C., and Geleta, M. (2022). Genetic diversity and population structure of sorghum [Sorghum bicolor (L.) moench] accessions as revealed by single nucleotide polymorphism markers. *Front. Plant Sci.* 0, 3110. doi: 10.3389/fpls.2021.799482

Faye, J. M., Maina, F., Hu, Z., Fonceka, D., Cisse, N., and Morris, G. P. (2019). Genomic signatures of adaptation to sahelian and soudanian climates in sorghum landraces of senegal. *Ecol. Evol.* 9 (10), 6038–6051. doi: 10.1002/ece3.5187

Francis, R. M. (2017). Pophelper: an r package and web app to analyse and visualize population structure. *Mol. Ecol. Resour.* 17 (1). doi: 10.1111/1755-0998.12509

Gautier, M., Klassmann, A., and Vitalis, R. (2017). rehh 2.0: a reimplementation of the r package rehh to detect positive selection from haplotype structure. *Mol. Ecol. Resour* 17, 78–90. doi: 10.1111/1755-0998.12634

Harlan, J. R., and Stemler, A. (2012). "The races of sorghum in Africa," in *Origins of African plant domestication*, De Gruyter Mouton. 465–478. doi: 10.1515/9783110806373.465

Harlan, J. R., and Wet, J. M. J. (1972). A simplified classification of cultivated Sorghum1. *Crop Sci.* 12, 172–176. de. doi: 10.2135/cropsci1972.0011183X001200020005x

Haubold, B. (2014). Alignment-free phylogenetics and population genetics. *Briefings Bioinf.* 15, 407–418. doi: 10.1093/bib/bbt083

Hostetler, A. N., Govindarajulu, R., and Hawkins, J. S. (2021). QTL mapping in an interspecific sorghum population uncovers candidate regulators of salinity tolerance. *Plant Stress* 2. doi: 10.1016/j.stress.2021.100024

Huang, X., Wei, X., Sang, T., Zhao, Q., Feng, Q., Zhao, Y., et al. (2010). Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat. Genet.* 42, 961–967. doi: 10.1038/ng.695

Hurgobin, B., and Edwards, D. (2017). SNP discovery using a pangenome: Has the single reference approach become obsolete? *Biol. (Basel)* 6, 21. doi: 10.3390/biology6010021

Jensen, S. E., Charles, J. R., Muleta, K., Bradbury, P. J., Casstevens, T., Deshpande, S. P., et al. (2020). A sorghum practical haplotype graph facilitates genome-wide imputation and cost-effective genomic prediction. *Plant Genome* 13 (1), e20009. doi: 10.1002/tpg2.20009

Kebbede, W. Y. (2020). Genetic variability and divergence in sorghum: Review. *Int. J. Res. Stud. Agric. Sci. (IJRSAS)* 6, 2454–6224. doi: 10.20431/2454-6224.0605002

Lam, H. M., Xu, X., Liu, X., Chen, W., Yang, G., Wong, F. L., et al. (2010). Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. *Nat. Genet.* 42, 1053–1059. doi: 10.1038/ng.715

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923

Letunic, I., and Bork, P. (2019). Interactive tree of life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Res.* 47 (W1), W256–W259. doi: 10.1093/nar/gkz239

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

Li, L. F., Li, Y. L., Jia, Y., Caicedo, A. L., and Olsen, K. M. (2017). Signatures of adaptation in the weedy rice genome. *Nat. Genet.* 49(5), 811–814. doi: 10.1038/ng.3825

Lin, Y. L., Chang, P. C., Hsu, C., Hung, M. Z., Chien, Y. H., Hwu, W. L., et al. (2022). Comparison of GATK and DeepVariant by trio sequencing. *Sci. Rep.* 12 (1). doi: 10.1038/s41598-022-05833-4

Mace, E. S., Tai, S., Gilding, E. K., Li, Y., Prentis, P. J., Bian, L., et al. (2013). Whole-genome sequencing reveals untapped genetic potential in africa's indigenous cereal crop sorghum. *Nat. Commun.* 4, 2320. doi: 10.1038/ncomms3320

Mapleson, D., Accinelli, G. G., Kettleborough, G., Wright, J., and Clavijo, B. J. (2017). KAT: a K-mer analysis toolkit to quality control NGS datasets and genome assemblies. *Bioinformatics* 33, 574. doi: 10.1093/bioinformatics/btw663

Marçais, G., and Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* 27, 764–770. doi: 10.1093/bioinformatics/btr011

Massel, K., Campbell, B. C., Mace, E. S., Tai, S., Tao, Y. , Worland, B. G., et al. (2016). Whole genome sequencing reveals potential new targetsfor improving nitrogen uptakeand utilization in sorghumbicolor sorghumbicolor. *Front. Plant Sci.* 7, 1544. doi: 10.3389/fpls.2016.01544

Mccouch, S., Baute, G. J., Bradeen, J., Bramel, P., Bretting, P. K., Buckler, E., et al. (2013). Agriculture: Feeding the future : Nature : Nature publishing group. *Nature* 499, 23–24. doi: 10.1038/499023a

McNally, K. L., Childs, K. L., Bohnert, R., Davidson, R. M., Zhao, K., Ulat, V. J., et al. (2009). Genomewide SNP variation reveals relationships among landraces and modern varieties of rice. *Proc. Natl. Acad. Sci.* 106, 12273–12278. doi: 10.1073/pnas.0900992106

Morris, G. P., Ramu, P., Deshpande, S. P., Hash, C. T., Shah, T., Upadhyaya, H. D., et al. (2013). Population genomic and genome-wide association studies of agroclimatic traits in sorghum. *Proc. Natl. Acad. Sci. U.S.A.* 110 (2), 453–458. doi: 10.1073/pnas.1215985110

Nazir, M. F., Jia, Y., Ahmed, H., He, S., Iqbal, M. S., Sarfraz, Z., et al. (2020). Genomic insight into differentiation and selection sweeps in the improvement of upland cotton. *Plants* 9 (6). doi: 10.3390/plants9060711

Nordström, K. J. V., Albani, M. C., James, G. V., Gutjahr, C., Hartwig, B., Turck, F., et al. (2013). Mutation identification by direct comparison of whole-genome sequencing data from mutant and wild-type individuals using k-mers. *Nat. Biotechnol.* 31, 325–330. 2013 31:4. doi: 10.1038/nbt.2515

Norman, A., Taylor, J., Edwards, J., and Kuchel, H. (2018). Optimising genomic selection in wheat: Effect of marker density, population size and population structure on prediction accuracy. *G3 Genes|Genomes|Genetics* 8, 2889–2899. doi: 10.1534/g3.118.200311

Ondov, B. D., Treangen, T. J., Melsted, P., Mallonee, A. B., Bergman, N. H., Koren, S., et al. (2016). Mash: Fast genome and metagenome distance estimation using MinHash. *Genome Biol.* 17 (1), 132. doi: 10.1186/s13059-016-0997-x

Pajuste, F. D., Kaplinski, L., Möls, M., Puurand, T., Lepamets, M., and Remm, M. (2017). FastGT: An alignment-free method for calling common SNVs directly from raw sequencing reads. *Sci. Rep.* 7 (1). doi: 10.1038/s41598-017-02487-5

Paradis, E., Claude, J., and Strimmer, K. (2004). APE: Analyses of phylogenetics and evolution in r language. *Bioinformatics* 20, 289–290. doi: 10.1093/bioinformatics/btg412

Paterson, A. H., Bowers, J. E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., et al. (2009). The sorghum bicolor genome and the diversification of grasses. *Nature* 457, 551–556. doi: 10.1038/nature07723

Peng, Y., Leung, H. C. M., Yiu, S. M., and Chin, F. Y. L. (2012). IDBA-UD: A de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 28, 1420–1428. doi: 10.1093/bioinformatics/bts174

Poplin, R., Chang, P. C., Alexander, D., Schwartz, S., Colthurst, T., Ku, A., et al. (2018). A universal SNP and small-indel variant caller using deep neural networks. *Nat. Biotechnol.* 36, 983–987. doi: 10.1038/nbt.4235

Ramu, P., Billot, C., Rami, J. F., Senthilvel, S., Upadhyaya, H. D., Ananda Reddy, L., et al. (2013). Assessment of genetic diversity in the sorghum reference set using EST-SSR markers. *Theor. Appl. Genet.* 126 (8). doi: 10.1007/s00122-013-2117-6

Ruperao, P., Thirunavukkarasu, N., Gandham, P., Selvanayagam, S., Govindaraj, M., Nebie, B., et al. (2021). Sorghum pan-genome explores the functional utility for genomic-assisted breeding to accelerate the genetic gain. *Front. Plant Sci.* 12, 963. doi: 10.3389/fpls.2021.666342

Sabeti, P. C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., et al. (2007). Genome-wide detection and characterization of positive selection in human populations. *Nature* 449, 913. doi: 10.1038/nature06250

Saleem, A., Muylle, H., Aper, J., Ruttink, T., Wang, J., Yu, D., et al. (2021). A genome-wide genetic diversity scan reveals multiple signatures of selection in a European soybean collection compared to Chinese collections of wild and cultivated soybean accessions. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.631767

Sapkota, S., Boyles, R., Cooper, E., Brenton, Z., Myers, M., and Kresovich, S. (2020). Impact of sorghum racial structure and diversity on genomic prediction of grain yield components. *Crop Sci.* 60, 132–148. doi: 10.1002/csc2.20060

Saxena, R. K., Edwards, D., and Varshney, R. K. (2014). Structural variations in plant genomes. *Briefings Funct. Genomics* 13, 296–307. doi: 10.1093/bfgp/elu016

Sheppard, S. K., Didelot, X., Meric, G., Torralbo, A., Jolley, K. A., Kelly, D. J., et al. (2013). Genome-wide association study identifies vitamin B5 biosynthesis as a host specificity factor in campylobacter. *Proc. Natl. Acad. Sci. U.S.A.* 110, 11923–11927. doi: 10.1073/pnas.1305559110

Simpson, J. T., Wong, K., Jackman, S. D., Schein, J. E., Jones, S. J. M., and Birol, I. (2009). ABySS: A parallel assembler for short read sequence data. *Genome Res.* 19, 1117–1123. doi: 10.1101/gr.089532.108

Smýkal, P., Nelson, M. N., Berger, J. D., and von Wettberg, E. J. B. (2018). The impact of genetic changes during crop domestication *Agronomy* 2018, vol. 8 Page 119, 8, 119. doi: 10.3390/agronomy8070119

Song, K., Ren, J., Reinert, G., Deng, M., Waterman, M. S., and Sun, F. (2014). New developments of alignment-free sequence comparison: measures, statistics and next-generation sequencing. *Briefings Bioinf.* 15, 343–353. doi: 10.1093/bib/bbt067

Stephan, W., Wiehe, T. H. E., and Lenz, M. W. (1992). The effect of strongly selected substitutions on neutral polymorphism: Analytical results based on diffusion theory. *Theor. Population Biol.* 41, 237–254. doi: 10.1016/0040-5809(92)90045-U

Swarts, K., Gutaker, R. M., Benz, B., Blake, M., Bukowski, R., Holland, J., et al. (2017). Genomic estimation of complex traits reveals ancient maize adaptation to temperate north america. *Science* 1979), 357. doi: 10.1126/science.aam9425

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123(3), 585–595. doi: 10.1093/genetics/123.3.585

Tao, Y., Luo, H., Xu, J., Cruickshank, A., Zhao, X., Teng, F., et al. (2021). Extensive variation within the pan-genome of cultivated and wild sorghum. *Nat. Plants* 7, 766–773. doi: 10.1038/s41477-021-00925-x

Telenti, A., Lippert, C., Chang, P. C., and DePristo, M. (2018). Deep learning of genomic variation and regulatory network data. *Hum. Mol. Genet.* 27, R63–R71. doi: 10.1093/hmg/ddy115

Tenaillon, M. I., Sawkins, M. C., Long, A. D., Gaut, R. L., Doebley, J. F., and Gaut, B. S. (2001). Patterns of DNA sequence polymorphism along chromosome 1 of maize (Zea mays ssp. *mays L.). Proc. Natl. Acad. Sci.* 98, 9161–9166. doi: 10.1073/pnas.151244298

Valluru, R., Gazave, E. E., Fernandes, S. B., Ferguson, J. N., Lozano, R., Hirannaiah, P., et al. (2019). Deleterious mutation burden and its association with complex traits in sorghum (Sorghum bicolor). *Genetics* 211 (3), 1075–1087. doi: 10.1534/genetics.118.301742

Vatsiou, A. I., Bazin, E., and Gaggiotti, O. E. (2016). Detection of selective sweeps in structured populations: A comparison of recent methods. *Mol. Ecol.* 25 (1). doi: 10.1111/mec.13360

Voight, B. F., Kudaravalli, S., Wen, X., and Pritchard, J. K. (2006). A map of recent positive selection in the human genome. *PloS Biol.* 4, 0446–0458. doi: 10.1371/journal.pbio.0040072

Wang, Y., Fu, L., Ren, J., Yu, Z., Chen, T., and Sun, F. (2018). Identifying group-specific sequences for microbial communities using long k-mer sequence signatures. *Front. Microbiol.* 9, 872. doi: 10.3389/fmicb.2018.00872

Wang, Y., Tan, L., Fu, Y., Zhu, Z., Liu, F., Sun, C., et al. (2015). Molecular evolution of the sorghum maturity gene Ma3. *PloS One* 10 (5). doi: 10.1371/journal.pone.0124435

Wang, Y., Lei, X., Wang, S., Wang, Z., Song, N., Zeng, F., et al. (2016). Effect of k-tuple length on sample-comparison with high-throughput sequencing data. *Biochem. Biophys. Res. Commun.* 469, 1021–1027. doi: 10.1016/j.bbrc.2015.11.094

Weir, B. S., and Cockerham, C. C. (1984). ESTIMATING f-STATISTICS FOR THE ANALYSIS OF POPULATION STRUCTURE. *Evol. (N Y)* 38, 1358–1370. doi: 10.1111/j.1558-5646.1984.tb05657.x

Wendorf, F., Close, A. E., Schild, R., Wasylikowa, K., Housley, R. A., Harlan, J. R., et al. (1992). Saharan exploitation of plants 8,000 years BP. *Nature* 359, 721–724. doi: 10.1038/359721a0

Wilson, B. A., Pennings, P. S., and Petro, D. A. (2017). Soft selective sweeps in evolutionary rescue. *Genetics* 205 (4). doi: 10.1534/genetics.116.191478

Wu, S., Wang, X., Reddy, U., Sun, H., Bao, K., Gao, L., et al. (2019). Genome of 'Charleston gray', the principal american watermelon cultivar, and genetic characterization of 1,365 accessions in the U.S. *Natl. Plant Germplasm System watermelon collecton. Plant Biotechnol. J.* 17 (12), 2246–2258. doi: 10.1111/pbi.13136

Yan, S., Wang, L., Zhao, L., Wang, H., and Wang, D. (2018). Evaluation of genetic variation among sorghum varieties from southwest China via genome resequencing. *Plant Genome* 11, 170098. doi: 10.3835/plantgenome2017.11.0098

Zerbino, D. R., and Birney, E. (2008). Velvet: Algorithms for de novo short read assembly using de bruijn graphs. *Genome Res.* 18, 821–829. doi: 10.1101/gr.074492.107

Zhang, L. M., Leng, C. Y., Luo, H., Wu, X. Y., Liu, Z. Q., Zhang, Y. M., et al. (2018). Sweet sorghum originated through selection of dry, a plant-specific nac transcription factor gene[open]. *Plant Cell* 30 (10), 2286–2307. doi: 10.1105/tpc.18.00313

Zhong, L., Zhu, Y., and Olsen, K. M. (2022). Hard versus soft selective sweeps during domestication and improvement in soybean. *Mol. Ecol.* 31 (11), 3137–3153. doi: 10.1111/mec.16454

Zhou, Y., Zhao, X., Li, Y., Xu, J., Bi, A., Kang, L., et al. (2020). Triticum population sequencing provides insights into wheat adaptation. *Nat. Genet.* 52, 1412–1422. doi: 10.1038/s41588-020-00722-w

Zielezinski, A., Vinga, S., Almeida, J., and Karlowski, W. M.. (2017). Alignment-free sequence comparison: Benefits, applications, and tools. *Genome Biol.* 18 (1). doi: 10.1186/s13059-017-1319-7

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 15 May 2023
DOI 10.3389/fpls.2023.1169984

OPEN ACCESS

EDITED BY
Yinglong Chen,
University of Western Australia,
Australia

REVIEWED BY
Shuiming Xiao,
China Academy of Chinese Medical Sciences, China
Mohammad Nadeem,
King Saud University, Saudi Arabia

*CORRESPONDENCE
Madhvi Joshi
madhvimicrobio@gmail.com

RECEIVED 20 February 2023
ACCEPTED 17 April 2023
PUBLISHED 15 May 2023

CITATION
Travadi T, Shah AP, Pandit R, Sharma S, Joshi C and Joshi M (2023) A combined approach of DNA metabarcoding collectively enhances the detection efficiency of medicinal plants in single and polyherbal formulations.
Front. Plant Sci. 14:1169984.
doi: 10.3389/fpls.2023.1169984

A combined approach of DNA metabarcoding collectively enhances the detection efficiency of medicinal plants in single and polyherbal formulations

Tasnim Travadi, Abhi P. Shah, Ramesh Pandit, Sonal Sharma, Chaitanya Joshi and Madhvi Joshi*

Gujarat Biotechnology Research Centre (GBRC), Department of Science and Technology, Government of Gujarat, Gandhinagar, Gujarat, India

Introduction: Empirical research has refined traditional herbal medicinal systems. The traditional market is expanding globally, but inadequate regulatory guidelines, taxonomic knowledge, and resources are causing herbal product adulteration. With the widespread adoption of barcoding and next-generation sequencing, metabarcoding is emerging as a potential tool for detecting labeled and unlabeled plant species in herbal products.

Methods: This study validated newly designed *rbcL* and *ITS2* metabarcode primers for metabarcoding using in-house mock controls of medicinal plant gDNA pools and biomass pools. The applicability of the multi-barcode sequencing approach was evaluated on 17 single drugs and 15 polyherbal formulations procured from the Indian market.

Results: The *rbcL* metabarcode demonstrated 86.7% and 71.7% detection efficiencies in gDNA plant pools and biomass mock controls, respectively, while the *ITS2* metabarcode demonstrated 82.2% and 69.4%. In the gDNA plant pool and biomass pool mock controls, the cumulative detection efficiency increased by 100% and 90%, respectively. A 79% cumulative detection efficiency of both metabarcodes was observed in single drugs, while 76.3% was observed in polyherbal formulations. An average fidelity of 83.6% was observed for targeted plant species present within mock controls and in herbal formulations.

Discussion: In the present study, we achieved increasing cumulative detection efficiency by combining the high universality of the *rbcL* locus with the high-resolution power of the *ITS2* locus in medicinal plants, which shows applicability of multilocus strategies in metabarcoding as a potential tool for the Pharmacovigilance of labeled and unlabeled plant species in herbal formulations.

KEYWORDS

authentication, DNA metabarcoding, herbal medicines, next generation sequencing, pharmacovigilance

1 Introduction

The herbal commodity market is thriving, owing to the widespread belief that traditional medicine is natural and thus safer, thereby promoting good health and sustainable life policies. But as the market expands, the shortage of genuine resources and lack of taxonomic knowledge challenge the authenticity of herbal drugs and increase the incidence of economically motivated or unintentional adulterations and/or substitutions (Raclariu et al., 2018a). Strict pharmacovigilance is necessary to retain the trust and safety of consumers and their health. However, the regulatory guidelines for medicinal plants often blur the line between foods and therapeutics and vary from nation to nation. To close this research gap, regulatory bodies must implement more reliable, universal, and robust detection methods (Shetti et al., 2011).

Nowadays, various pharmacopeia advocate DNA-based methods such as DNA barcoding and species-specific PCR assay to authenticate herbal raw material. DNA is more stable, unaffected by external factors, and invariably present in almost all plant tissues (Wu and Shaw, 2022). In addition, the DNA-based results are independent of seasonal variations in the age of the plant, which in the case of chemical marker-based methods vary significantly (Parveen et al., 2016). Therefore, the results of DNA-based methods are free from subjectivity, accurate, and provide a universally accepted platform for the authentication of botanicals in a wide range of food and herbal products (Lo and Shaw, 2019). The advent of DNA barcoding is the first step in this direction, as barcoding gives resolution up to species level (Hebert et al., 2003). There are 17 potential barcode regions (*matK, rbcL, ITS, ITS2, psbA-trnH, atpF-atpH, ycf5, psbKI, psbM-trnD, coxI, nad1, trnL-F, rpoB, rpoC1, atpF-atpH*, and *rps16*) for plants, having different degrees of universality, specificity, and taxa resolution power that extensively used in the authentication and identification of medicinal plants (Parveen et al., 2016; Kress, 2017). However, DNA barcoding (Vassou et al., 2016) and species-specific assays (Sharma and Shrivastava, 2016; Travadi et al., 2022a; Travadi et al., 2022b) cannot resolve presence of multiple plant species in a single sample (Mishra et al., 2016), that limitation could be overcome by DNA metabarcoding.

DNA metabarcoding combined the strengths of next-generation sequencing and barcoding for detecting multiple taxa in samples (Coghlan et al., 2012). Using a single plant barcode for species-level identification has proven challenging due to the great diversity, relatively slow molecular evolution, frequent cross-pollinations, and hybridization in the plant kingdom; henceforth, different barcodes show different degrees of taxon specificity (Fazekas et al., 2009). To precisely identify the plant species in the sample, multi-barcode approaches have become more prevalent. Xin et al. (2018) and Frigerio et al. (2021) employed a multi-barcode approach of *ITS2* and *trnL* for various Chinese medicine and herbal teas. However, these studies also highlighted the limitations of DNA metabarcoding applications for authentication of herbal products due to variability in degrees of universality and resolution power of barcodes for specific taxa, a lack of a curated database, and a robust bioinformatics pipeline. To overcome these constraints, there is a need for screening of new barcodes and new variable regions within the same barcode for authentication of the herbal products.

Therefore, the aims of the present study were: 1) to develop a new *rbcL* and *ITS* metabarcode for the detection of medicinal plant species, 2) to validate the primers specificity and efficiency using mock controls and 3) to see whether a multi-barcoding approach could be used for the pharmacovigilance of the herbal products? (17 different single plant formulation and 15 polyherbal market formulations in this study), and 3) to see whether a multi-barcoding approach could detect targeted plant species in herbal formulations?

2 Materials and methods

2.1 Collection of reference plant material and herbal products

Reference plant materials were collected with the aid of a taxonomist from the Maharaja Sayajirao University (MSU), Vadodara (Gujarat, India) and the Directorate of Medicinal and Aromatic Plants Research (DMAPR), Anand (Gujarat, India). As described earlier, reference plant materials were authenticated by Sanger sequencing of *rbcL* gene (Pandit et al., 2021), and sequences were submitted to the NCBI database (accession number MW628906 to MW628936). Voucher specimens were developed and deposited in our institutional herbarium.

32 herbal products were collected by blind sampling from the local market and e-commerce, with 17 single drugs and 15 polyherbal formulations. Single drugs include four Tulsi (*Ocimum tenuiflorum*) powders, five Gokhru (*Tribulus terrestris*) powders, three Shatavari (*Asparagus racemosus*) powders, two Vasa (*Justicia adhatoda*) products, and one each of Bhringraj (*Eclipta alba*), Ashwagandha (*Withania somnifera*), and Arjuna (*Terminalia arjuna*) powder. Polyherbal formulations include three market samples of Trikatu (has three plant species), three samples of Sitopladi [comprises five constituents, only three of which are plant species; the other two are sugar and Vanshlochan (the female bamboo exudate), hence these two constituents were not considered while analyzing the data expecting absence of DNA for these two], four samples of Rasayana (has three plant species), four samples of Hingwashtak (has seven plant species), and one sample of Talisadi [comprises eight constituents, only six of which are plant species; the other two are sugar and Vanshlochan (the female bamboo exudate), hence these two constituents were not considered while analyzing the data expecting absence of DNA for these two].

2.2 Primer designing

To design the metabarcodes for *ITS2* gene, *ITS2* sequences of Magnoliophyta from the BOLD database were downloaded and curated, particularly for the length. For the *rbcL* gene, we used 1,776 sequences of our *in-house* sequencing project submitted to the

BOLD database. To design universal barcodes, *rbcL* gene sequences were filtered by length between 450-600 bp. At the end, 1,465 and 1,701 *ITS2* and *rbcL* sequences were retained to design the metabarcode. These sequences were preceded for multiple sequence alignment separately (*ITS2* and *rbcL*) using BioEdit 7.2. HYDEN (HighlY DEgeNerate primers) software (Linhart and Shamir, 2005) to design degenerate primers, where the maximum of 3 degeneracy per primer was allowed. The designed primers were checked for amplicon length using NCBI primer BLAST (Altschul et al., 1990). *rbcL* reverse primer sequence was designed in this study, while forward primer sequence was obtained from Maloukh et al. (2017). To synthesize fusion primers, forward primers of *rbcL* and *ITS2* were tagged with the Ion torrent adapter and a ten bp multiplex identifier barcode. In contrast, reverse primers were tagged with the P1 adapter. Nucleotide sequence of the designed primers and their amplicon length are shown in Table 1.

2.3 PCR Optimization and library preparation

The library preparation process became a single-step process with barcoded fusion primers. The PCR optimization with each barcoded fusion primer was done with 45 different plant DNA listed in Supplementary Information Table S1. Thermal cycler conditions, especially primer annealing temperature, were optimized for *rbcL* and *ITS2* primer pairs with the following conditions. PCR mixture containing 10 µL Emerald Master mix (2X) (TaKaRa), 2 µL total genomic DNA (10-15 ng/µL), 1 µL of forward (5 pmol), 1 µL of reverse primer (5 pmol), 1 µL BSA (2 mg/mL) and 5 µL PCR grade water with the following thermal cycling conditions. Initial denaturation 95°C for 5 minutes, followed by 30 cycles of 95°C for 1 minute, for primer annealing a temperature gradient of 50°C to 60°C with an interval of 2°C for 30 seconds and 72°C for 1 minute, and final extension 72°C for 5 minutes. 2.4 Preparation of different mock controls

Three different types of controls were prepared as follows: Control 1) genomic DNA from plant leaves from different genera belonging to diverse families has been first isolated and pooled into three different groups as mentioned below, Control 2) simulated plant biomass controls (blended formulations) in which individual plant part having medicinal value has been mixed in three groups as control one and subjected to DNA isolation, and Control 3) genomic DNA (Isolated from plant leaves) pool from different species of the two genera (Figure 1). As mentioned above, three different groups were prepared for the first type of control with the plant species of a different genus. Group one comprised DNA of five species in equal proportion (5P), and further, in groups 2 and 3, DNA was added from ten (10P) and fifteen different plant species (15P) (Figure 1, 2A). High-quality DNA of all the species have been isolated individually from leaf tissue and pooled together in equal proportion to make these groups. The group's diversity has been increased by adding species from diverse genera that belong to diverse families to evaluate the resolution power and universality of the primers for the maximum number of species. For the second type of control, the same three groups of plants were used in the first controls (labeled as 5S, 10S, 15S). Still, simulated blended plant parts containing bioactive therapeutics were mixed in equal proportion (biomass admixture controls). These controls can be used to comprehend biases introduced during the DNA extraction and PCR dynamics under the influence of secondary metabolites on PCR amplification. For the third type of control, two groups were prepared. Group one comprises six plant species of the two genera, including *Asparagus* and *Terminalia* (Figure 2B). The second group includes seven plant species of the two genera, including *Piper* and *Phyllanthus* (Figure 2B). Similar to the first control, high-quality DNA was individually isolated from each species and pooled in equal amounts. These controls were utilized to obtain insight into the resolving strength of our newly designed *rbcL* and *ITS2* metabarcodes at lower taxa levels.

2.4 Metabarcoding

DNA from plant materials, blended formulations, and herbal products 'was extracted in duplicate using DNeasy Plant Mini Kit (QIAGEN, Germany) following themanufacturer's instructions. The library was prepared from each DNA sample using *rbcL* and *ITS2* fusion primers with optimized PCR conditions. The libraries were purified using AMPure XP beads (Beckman Coulter, CA, USA), and the quality of some of the libraries was checked using Agilent high-sensitivity DNA kit on Agilent 2100 Bioanalyzer. For each sample, libraries from two replicates were pooled. Further, all libraries were diluted to 100 pM and pooled in equimolar concentration. Emulsion PCR was carried out using Ion 520™ & Ion 530™ Kit-OT2 with 400 bp chemistry (Thermo Fisher Scientific, MA, USA). Sequencing was performed on the Ion S5 system using a 520/530 chip (Thermo Fisher Scientific, MA, USA). The data have been submitted to the National Centre for Biotechnology Information (NCBI) BioProject database under

TABLE 1 Primer sequences of *rbcL* and *ITS2* metabarcode with their annealing temperature.

Metabarcode	Primer	Sequence (5′ → 3′)	References	Annealing Temperature	Amplicon Length
rbcL	*rbcL-F*	ATGTCACCACAAACAGAGACTAAAGC	20	60°C	320-350 bp
	rbcL-R	GTARCVRAMCCTTCTTCAAAAAGGTC	This study		
ITS2	*ITS2*-F	CRRAATCCCGTGAACCATCGAGTCYT	This study	60°C	310-330 bp
	ITS2-R	AGCGGGTRRTCCCRCCTGACYTG	This study		

10.3389/fpls.2023.1169984

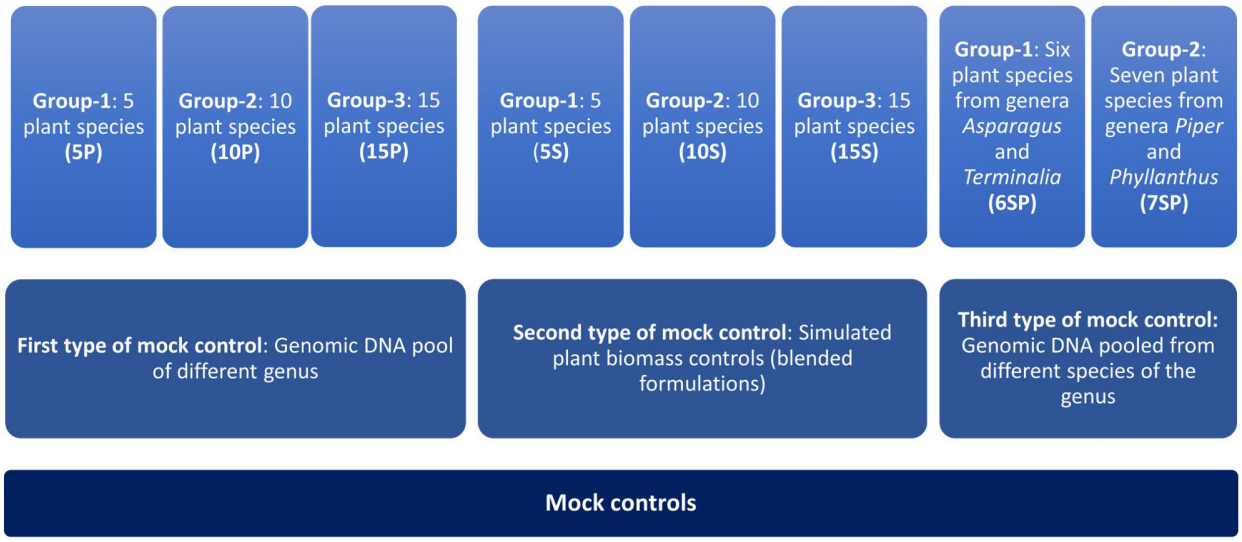


FIGURE 1
Schematic representation of different types of mock controls prepared in this study. First and second type of mock controls has 3 different groups comprising 5, 10, and 15 plant species. Third type of mock control has 2 different groups, one with a gDNA pool of six plant species from genera *Asparagus* and *Terminalia* and another with a gDNA pool of seven plant species from genera *Piper* and *Phyllanthus*.

A

Sr. no.	Plant Species	Family	Plant part used in blended mock control	Mock control ID#
1	*Andrographis paniculata*	Acanthaceae	Whole plant	5P/5S
2	*Azadirachcta indica*	Meliaceae	Leaves	
3	*Eclipta alba*	Asteraceae	Whole plant	
4	*Piper nigrum*	Piperaceae	Seed	
5	*Zingiber officinale*	Zingiberaceae	Rhizome	10P/10S
6	*Aegale marmelous*	Rutaceae	Fruits	
7	*Centella asiatica*	Apiaceae	Whole plant	
8	***Terminalia arjuna***	Combretaceae	Bark	
9	***Terminalia bellerica***	Combretaceae	Fruits	
10	*Vitex negundo*	Lamiaceae	Leaves	
11	*Bacopa monneri*	Plantaginaceae	Whole plant	15P/15S
12	*Sena tora*	Fabaceae	Leaves	
13	*Phyllanthus embilica*	Phyllanthaceae	Fruits	
14	***Terminalia chebula***	Combretaceae	Fruits	
15	*Justicia adhatoda*	Acanthaceae	Leaves	

Detected and not detected plants in metabarcoding

5P			5S			10P			10S			15P			15S		
rbcL	*ITS2*	C	*rbcL*	*ITS2*	C	*rbcL*	*ITS2*	C	*rbcL*	*ITS2*	C	*rbcL*	*ITS2*	C	*rbcL*	*ITS2*	C

B

Sr.no	Plant species	Family	Mock control ID	*rbcL*	*ITS2*	C
1	*Asparagua racemosus*	Asparagaceae	6SP			
2	*Asparagua dumosus*					
3	*Asparagus adscendens*					
4	***Terminalia arjuna***	Combretaceae				
5	***Terminalia bellerica***					
6	***Terminalia chebula***					
1	*Pipper nigrum*	Piperaceae	7SP			
2	*Pipper logume*					
3	*Piper betle*					
4	*Pipper arabarum*					
5	*Phyllanthus embica*	Phyllanthaceae				
6	*Phyllanthus amaras*					
7	*Phyllanthus madresapatensis*					

Detected
Not-detected

FIGURE 2
The distribution of predefined herbal species detected in each mock control with *rbcL*, *ITS2*, and combined metabarcoding approach. **(A)** Detected and undetected plant species with *rbcL*, *ITS2*, and combined approach in type one mock control, i.e., genomic DNA pool of different genera, and type two mock control, i.e., simulated plant biomass control (blended formulations) comprising three groups having five plant species (5P/5S), ten plant species (10P/10S) and 15 plant species (15P/15S). **(B)** Detected and undetected plant species with *rbcL*, *ITS2*, and combined approach in a third type of mock controls (i.e., genomic DNA pooled from different species of the genus) comprising two different groups, one having six plant species (6SP) from genera *Asparagus* and *Terminalia* and another having seven plant species (7SP) from genera *Piper* and *Phyllanthus*. Detected plant species are represented in pink, while undetected plant species are represented in grey. Plant species, their family, plant parts used in type two mock controls (simulated plant biomass controls or blended formulations), and mock control IDs are also indicated in the figure. P: gDNA plant pool controls, S: simulated blended plant pool (biomass control), C: a combined approach.

accession number PRJNA960808 (https://www.ncbi.nlm.nih.gov/bioproject/).

2.5 Metabarcoding data analysis

Optimization was done with three parameters for establishing the metabarcoding data analysis pipeline. 1) filtering criteria includes discarding reads with length <280 and >350 bp, <300 and >350 bp, <320 and >350 bp for *rbcL* and for *ITS2* discarding reads <280 and >300 bp; 2) OTU clustering with 97, 98 and 99% similarity in reads; 3) discarding OTU clusters having <5 and <10 reads. Obtained reads were filtered based on the quality score (Q >=25 for *rbcL* and Q >=20 for *ITS*) and read length using PRINSEQ (Schmieder and Edwards, 2011). Clustering filtered reads were performed using CD-HIT-EST (Huang et al., 2010). After that, the taxonomic assignment of OTU clusters having ≥5 and ≥10 reads was done using BLASTn (Altschul et al., 1990) (NCBI) with minimum E value $10E^{-5}$. For each sequence, ten hits were retrieved, and each hit was inspected and evaluated manually for the assigned plant genus and species. To analyze read abundance of each plant species, the number of reads was normalized by considering a total number of reads obtained after discarding clusters with <5 reads and <10 reads as 100%. The following formula was calculated for detection efficiency (%) for both metabarcodes. (Total number of detected targeted or labeled plant species/total number of plant species present in herbal formulation) × 100. Fidelity of detection (absolute) can be defined as the total number of samples or herbal formulations in which targeted plant species were detected per the total number of samples or herbal formulations in which targeted plant species were present (Seethapathy et al., 2019). The relative fidelity of detection (%) was calculated using the following formula. (Total number of samples or herbal formulations in which targeted plant species detected/total number of samples or herbal formulations in which targeted plant species are present) × 100. Fidelity of detection (absolute or relative) was calculated only where plant species that present in more than one group of the same type of control (n>1) and for market samples, sample size is more than one (n>1).

3 Results and discussion

Coghlan et al. (2012) first introduced a metabarcoding approach for detecting plant and animal raw materials used in 15 traditional Chinese medicines using a P-loop region of the plastid *trnL* gene and 16S mtDNA marker, respectively. Later, several studies reported application of metabarcoding for authentication and detection of plant materials in herbal medicines with single and multi-barcode approaches. For instance, Yao et al. (2022) and Cheng et al. (2014) employed a multi-barcode approach of *ITS2* and *trnL* in metabarcoding for detection of plant species in various traditional Chinese medicine (TCM). Urumarudappa et al. (2020) used *ITS2* and *rbcL* barcode in metabarcoding for detection of plant species in herbal medicines of Thailand. Using *ITS1* and *ITS2* barcodes in metabarcoding, Raclariu et al. (2017a; 2017b; 2018b) reported presence of unlabeled species by 89, 68, and 15% in single drugs of *Echinacea* species, *Hypericum perforatum*, and *Veronica officinalis*, respectively sold in the European market.

In 2014 and 2015, the total commercial market for herbal materials in India was estimated to be more than 512,000 tonnes, with a market value of USD 1 billion (Ved and Goraya, 2007). India has over 8,000 authorized medicinal product manufacturing units, and the market growth for herbal products is outstripping supply capacity for some plant species (Ved and Goraya, 2007). However, to date, detection of raw plant materials of Indian-marketed herbal medicine using metabarcoding approach is not well established. Ichim (2019) demonstrated 31% adulteration in 752 Indian-marketed herbal products with DNA barcoding and species-specific marker-based approach but not *via* metabarcoding. Earlier, we reported presence of unspecified plant species in four polyherbal formulations of the Indian market using *rbcL* minibarcode *via* metabarcoding approach (Pandit et al., 2021). Here, we have used a multi-barcode strategy to identify raw plant components in single drugs and polyherbal formulations of the Indian market using newly designed *rbcL* and *ITS2* metabarcodes.

3.1 PCR assays using newly designed *rbcL* and *ITS2* primers and fusion primers

Minimal criteria, such as universal amplifiabilities and minimum intraspecific but maximum interspecific divergence at the taxon level, must be followed in the search for the appropriate barcode region. Hence, degenerated *rbcL* and *ITS2* metabarcode primers were designed for high amplification efficiency, universality, and resolution power. In total, 45 medicinal plants from diverse families, genera, and species were taken to confirm and optimize the newly designed *rbcL* and *ITS2* primer sets for PCR amplification experimentally (Supplementary Information Table S1). The annealing temperature was optimized, and the results showed that the *rbcL* and *ITS2* primer sets performed optimum at 56°C (data not shown). Among newly designed *rbcL* and *ITS2* metabarcodes, *rbcL* is very robust and universal and gives a 100% amplification efficiency within selected 45 plants, but *ITS2* metabarcode gives 88.9% amplification efficiency and is not able to provide amplification in 5 plant species (11.1%) include *Ailanthus excelsa*, *Andrographis paniculata, A. vasica, O. tenuiflorum*, and *Ocimum canum* (Supplementary Information Table S1). However, due to the greater species-level discrimination power of *ITS2* in medicinal plants (Newmaster et al., 2013; Yao et al., 2022), *ITS2* metabarcode was taken together with *rbcL* metabarcode. The amplification effectiveness of "fusion primers" (tagged with Ion torrent adapter and barcodes) of *rbcL* and *ITS2* metabarcodes remained unchanged. However, the appearance of non-specific amplification in some barcodes suggests that 56°C annealing temperature is not optimal for the fusion primers. Therefore, further optimization of annealing temperature revealed that non-specific amplification was overcome by increasing the annealing temperature to 60°C (data not shown). At 60 °annealing temperature the amplification efficiency remained unaffected.

10.3389/fpls.2023.1169984

3.2 Establishing data analysis pipeline using mock controls

The first type of mock control, i.e., gDNA pooled controls of a different genus, was used to establish the metabarcoding data analysis pipeline. The first parameter is filtering criteria; for *ITS2*, the best filtering criteria was to remove reads with length < 300 bp, and for *rbcL*, discarding reads with length <300 and >350 bp (data not shown here). The second parameter is percentage similarity for the reads clustering (OTU picking), where in case of *rbcL* metabarcode, a greater number of plant species was detected when the reads were clustered at 99% identity. In the case of *ITS2* metabarcode, clustering the reads at 97% and 98% similarity were equally capable of resolving the plant species (Supplementary Information Figure S1). Therefore, for *ITS2* metabarcode, subsequently, for OTU clustering, i.e., 98% with a greater percentage was selected. The third parameter is the discarding OTU clusters having <5 or <10 reads, in which we observed that discarding OTU clusters having <5 reads was able to detect a greater number of plant species in the case of both metabarcodes (Supplementary Information Figure S2). Based on these findings, while selected read lengths were between 300 to 350 bp, 99% OTU clustering, and discarding of OTU clusters having <5 reads for the *rbcL* metabarcode and read lengths of >300 bp, 98% OTU clustering, and discarding of OTU clusters comprising <5 reads for the *ITS2* metabarcode for analyzing the metabarcoding data of other mock controls and commercial herbal formulations.

3.3 Metabarcoding of different types of mock controls

Total reads obtained after filtering and percentage of reads analyzed (from filtered reads) after discarding OTU clusters having <5 reads for each mock control are shown in Supplementary Information Table S2 for the first type of control, which is gDNA pooled controls of different genera comprising 5 (5P), 10 (10P), and 15 (15P) plant species total of 18657 and 452380 reads were obtained for *rbcL* and the *ITS2*, respectively (Supplementary Information Table S2). *Zingiber officinale* had the highest percentage of reads with *rbcL* metabarcode in 5P (45.9%) and 10P (30.9%), whereas *Senna tora* had the highest percentage (21.3%) in 15P (Supplementary Information Figure S3A). *Eclipta prostrata* had the highest percentage of reads with *ITS2* metabarcode in 5P (97.02%) and 10P (79.5%), whereas *Phyllanthus emblica* had the highest percentage (40.7%) in 15P (Supplementary Information Figure S3A). In 5P, out of five total four target plants were detected with *rbcL* metabarcode, and three were detected with *ITS2*. In 10P, out of ten, nine targeted plants were detected with *rbcL*, and seven targeted plants were detected with *ITS2*. In 15P, out of fifteen, thirteen targeted plants were detected with *rbcL*, and twelve targeted plants were detected with *ITS2* (Figure 2; Supplementary Information Figure S3A). On the whole, for the gDNA pooled controls of a different genus, detection efficiency of *rbcL* was observed at 80% for 5P and 10P, 86.7% for 15P. While detection efficiency of *ITS2* was observed at 80%, and combined detection efficiency of both metabarcodes was observed at 100% for all three gDNA pooled mock controls (Figures 2A, 3). Five plant species in all three gDNA pooled mock controls had 80% average fidelity, and the other five were present in two groups, i.e., 10P and 15P, had 90% average fidelity with *rbcL* metabarcode (Table 2). *ITS2* metabarcode exhibited 66.7% average fidelity for five plant species present in all three gDNA pooled mock controls and 90% average fidelity for other five plant species present in 10P and 15P. Combined average fidelity with both barcodes was 100% for gDNA pooled control (Table 2).

For the simulated plant biomass controls (i.e., blended formulations or second type of mock controls), 42140 and 334017

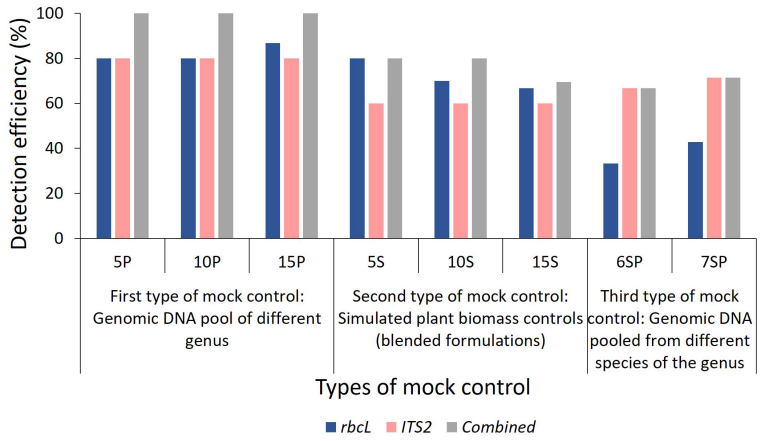


FIGURE 3
Detection efficiency obtained in mock controls by *rbcL*, *ITS2*, and combined approach. Detection efficiency (%) was calculated using a formula described in the Metabarcoding data analysis section of Materials and Methods. 5P: genomic DNA pool of five plant species, 10P: genomic DNA pool of ten plant species, 15P: genomic DNA pool of fifteen plant species, 5S: simulated blended plant pool of five plant species, 10S: simulated blended plant pool of ten plant species, 15S: simulated blended plant pool of fifteen plant species, 6SP: genomic DNA pool of six plant species from genera *Asparagus* and *Terminalia*, 7SP: genomic DNA pool of seven plant species from genera *Piper* and *Phyllanthus*. The detailed list of plant species used in each mock control is described in Figure 2.

TABLE 2 Detection fidelity in mock controls[≠].

Sr. no.	Plant Species	gDNA pooled mock controls				Simulated plant biomass controls (Blended formulations)			
			Relative fidelity/species (%)				Relative Fidelity/species (%)		
		Species present in no. of mock controls[#]	*rbcL*	*ITS2*	Combined	Species present in no. of mock controls[®]	*rbcL*	*ITS2*	Combined
1	*Andrographis paniculata*	3	100	0	100	3	100	0	100
2	*Azadirachcta indica*		100	33.3	100		100	66.7	100
3	*Eclipta alba*		100	100	100		100	100	100
4	*Piper nigrum*		0	100	100		100	100	100
5	*Zingiber officinale*		100	100	100		0	0	0
Average Fidelity (%)			**80**	**66.7**	**100**		**80**	**53.3**	**80**
6	*Aegale marmelous*	2	100	50	100	2	100	50	100
7	*Centella asiatica*		100	100	100		100	100	100
8	*Terminalia arjuna*		50	100	100		0	0	0
9	*Terminalia bellerica*		100	100	100		50	100	100
10	*Vitex negundo*		100	100	100		100	100	100
Average Fidelity (%)			**90**	**90**	**100**		**70**	**70**	**80**

[#]3: Species present in all three groups of gDNA pooled mock control (n=3); 2: Species present in two groups of gDNA pooled mock control i.e., 5P and 10P (n=2). [®]3: Species present in all three groups of simulated plant biomass mock control (n=3); 2: Species present in two groups of simulated plant biomass mock control i.e., 5S and 10S (n=2). [#®]Fidelity was not calculated for five plant species that present only in 15P and15S (n=1). [≠] Fidelity for the third type of mock control (i.e., gDNA pooled from different species of the genus) was not calculated as each plant species present only in one group (n=1).

reads were obtained after filtering for *rbcL* and the *ITS2*, respectively (Supplementary Information Table S2). The highest percentage of reads was observed for *Azadirachta indica* in 5S (62.5%) and 10S (34.7%), whereas, in 15S *Justicia adhatoda* (24.8%) showed highest percentage of reads using *rbcL* metabarcode (Supplementary Information Figure S3B). While in the case of *ITS2*, *Eclipta prostrata* had the highest percentage of reads in all simulated plant biomass controls (Supplementary Information Figure S3B). In 5S, 10S, and 15S total of 3, 7, and 10 targeted plants were detected by *rbcL*, and a total of 3, 6, and 8 targeted plants were detected by *ITS2*, respectively (Figure 2A; Supplementary Information Figure S3B). Detection efficiency for simulated plant biomass controls was observed 80% for 5S, 70% for 10S, and 66.7% for 15S with *rbcL* metabarcode. While detection efficiency of *ITS2* was observed at 60% for all three groups of simulated plant biomass control, and combined detection efficiency of both metabarcodes was observed 80% for 5S and 10S and 69.4% for 15S (Figure 3). Five plant species that were present in all three groups of simulated plant biomass control had 80% average fidelity, and other five plant species that were present in 10S and 15S had 70% average fidelity with *rbcL* metabarcode (Table 2). *ITS2* metabarcode exhibited 53.3% average fidelity for five plant species in all three groups and 70% average fidelity for five other plants present in two groups, i.e., 10S and 15S. Combined average fidelity with both barcodes was observed at 80% for simulated plant biomass controls (Table 2).

For the third type of control, gDNA pooled from different species of two genera, 10265 and 164358 reads were obtained for *rbcL* and the *ITS2*, respectively (Supplementary Information Table S2). In 6SP, out of six plant species total of three plant species include *Asparagus racemosus* (81.1%), *Terminalia bellirica* (11.8%), and *Terminalia chebula* (6.8%), were detected by *rbcL* while *ITS2* metabarcode was able to resolve five plant species except for *Asparagus dumosus* (Figure 2B; Supplementary Information Figure S3C). In 7SP, out of seven plant species, two plant species include *Piper nigrum* (24.7%) *and Phyllanthus emblica* (72.3%) were detected by rbcL. *In comparison,* ITS2 metabarcode was able to resolve four plant species, including *Piper nigrum* (0.8%), *Piper longum* (21%), *Phyllanthus emblica* (62.5%), *Phyllanthus amaras* (13.5%) (Figure 2B; Supplementary Information Figure S3C). *rbcL* showed 33.3%, and *ITS2* showed 66.7% detection efficiency in species-level control with a combined detection efficiency of 66.7% (Figures 2B, 2). These two species-level controls indicate a greater resolution spectrum of *ITS2* metabarcode than *rbcL* metabarcode. This finding corroborates with earlier reports in which authors demonstrated that *ITS2* metabarcode has greater species-level discrimination power than *rbcL* while *rbcL* has greater universality (Newmaster et al., 2013; Yao et al., 2022).

Here, in the first and third types of mock controls, DNA was pooled in equal proportions, and for the second type of simulated plant biomass controls, the equivalent weight of each plant species

part with therapeutic importance was mixed. The primary aim of all three mock controls was to evaluate the read abundance, detection efficiency, and fidelity differences introduced under the influence of secondary metabolites, primer fit compatibilities, and PCR dynamics. The impact was observed with percentage read variation of the same plant in a different control (Arulandhu et al., 2017). *Terminalia arjuna*, *T. chebula*, and *Phyllanthus emblica* were not detected in simulated plant biomass control might be due to variability in quality and quantity of DNA extracted from each plant species from the mixtures as different parts of plants, i.e., rhizome, fruits, leaves, and bark have been added into plant biomass controls (Ivanova et al., 2016). In addition to that, *ITS2* metabarcode is unable to resolve *A. paniculata* and *J. adhatoda* in all mock controls because the newly designed *ITS2* metabarcode is impotent in amplifying target sequence from these two plant species (Figure 1; Supplementary Information Table S1). Despite the mentioned limitations, the combined *rbcL* and *ITS2* metabarcoding approach could resolve plant species with high fidelity (Table 2) and can be implemented to detect plant species in herbal products.

3.4 Metabarcoding of single drugs

Tulsi (*O. tenuiflorum*) (The Ayurvedic Pharmacopoeia of India, Vol. II, 1999), Gokhru (*Tribulus terrestris*) (The Ayurvedic Pharmacopoeia of India, Vol. I,1990) Shatavari (*Asparagus racemosus*) (The Ayurvedic Pharmacopoeia of India, Vol. III, 2001), Vasa (*Justicia adhatoda*), Ashwagandha (*Withania somnifera*), Bhringraj (*Eclipta alba*), and Arjuna (*Terminalia arjuna*) were among the 17 single drugs that were collected. Total reads, reads obtained after filtering, and percentage of analyzed reads (from filtered reads) after discarding OTU clusters having <5 reads for each single drug are shown in Supplementary Information Table S3. For all single drugs, 135505 total raw reads for *rbcL* and 1390098 total raw reads for *ITS2* metabarcode were obtained (Supplementary Information Table S3). On an average, 12.6% (0-99.8%) reads have been obtained for non-targeted plant species with *rbcL* metabarcode and 46.5% (0-100%) reads have been obtained for non-targeted plant species with *ITS2* metabarcode in single drugs (Supplementary Information Table S4). Cross-contamination with allied species, harvesting process, pollen contamination, misidentification due to cryptic taxonomy, polynomial vernacular identification, manufacturing and packing procedure may contribute to the presence of non-targeted plant species (Seethapathy et al., 2019).

In tulsi powder (labeled as *O.tenuiflorum*), *rbcL* detected *O.tenuiflorum* ranging between 99.8% to 64.4%. Along with *O. tenuiflorum*, substituted *Ocimum basilicum* (Travadi et al., 2022b) was also observed in three herbal products with 17.7, 14.4, and 1.4% reads with *rbcL* metabarcode, respectively (Figure 4A). *O. tenuiflorum* could not be detected by *ITS2* metabarcode in any samples. However, *O. basilicum* was found in one sample with 11.6% reads (Figure 4B). That could be due to inability of new *ITS2* metabarcode to amplify *O. tenuiflorum* (Supplementary Information Table S1), which is proof of PCR biases toward the unintentional and low level of contamination present in samples and leads to the high number of reads for non-targeted plant species.

In Gokhru powder, *T.terrestris* (Chota Gokhru) was detected up to species level by *rbcL* with 100% reads in sample 5, 92.7% in sample 6, 0.2% in sample 7, 97.4% in sample 8, and 60.3% in sample 9 (Figure 4A). The *ITS2* metabarcode resolved *T. terrestris* only at the genus level with 100% in sample 5, 99% in sample 6, 15.5% in sample 7, 99.6% in sample 8, and 88.4% in sample 9. *ITS2* revealed the presence of 61% of *Pedalium murex* (Bada Gokhru) in sample 7 (Figure 4B), which is a commonly substituted species and labeled as Gokhru in Indian marketed herbal products (Choudhary et al., 2021).

In all three Shatavari powders, *rbcL* metabarcode detected *Asparagus setaceus* instead of *A. racemosus* with 98.3% reads in sample 10, 99.7% in sample 11, and 99.6% reads in sample 12. By *ITS2*, 67.7% to 61.7% reads for *A. racemosus* were obtained in all three Shatavari powders. In vasa powder, *J.adhatoda* was detected by *rbcL* with 84.2% reads in sample 13 and 99.4% in sample 14. In comparison, *ITS2* could not resolve *J.adhatoda* because of the amplification inability of our *ITS2* metabarcode (Figure 4, Supplementary Information Table S1). This result was consistent with the mock control 15P and 15S. In Ashwagandha powder (sample 16), *Withania coagulans* (77.9% reads) instead of *W. somnifera* was detected by *rbcL*. While with *ITS2*, *W. somnifera* was detected with 47.2% reads.

In Bhringraj powder, both *rbcL* and *ITS2* metabarcode detected *E.alba* with 98.8% and 95.7% reads, respectively. In Arjuna powder, *rbcL* metabarcode identified *Terminalia arjuna* only up to genus level with 95.5% reads, while *ITS2* metabarcode was identified *T. arjuna* at species level with 97.6% reads (Figure 4).

For Tulsi, Gokhru, Shatavari, and Vasa powder, 100% fidelity was obtained with *rbcL* metabarcode, while *ITS2* metabarcode exhibited 0% fidelity for Tulsi and Vasa powder and 100% fidelity for Gokhru and Shatavari powder (Table 3A). Suggesting that, for Tulsi and Vasa powder authentication, our *rbcL* metabarcode works efficiently but not *ITS2*. While *rbcL* metabarcode was not suitable for identifying P. murex in Gokhru powder, *rbcL* sequence for *Pedalium murex* was unavailable in NCBI (database was accessed on December 15, 2022). In addition, *T. arjuna* was resolved only up to genus level by *rbcL*, and *T. terrestris* was resolved up to only at the genus level by *ITS2* metabarcode. This observation is in concordance with mock controls and could be due to low interspecific variability of barcode sequence covered by our metabarcodes for resolving species to be identified. Although, combined metabarcoding approach provides 100% detection efficiency with 100% fidelity for single drugs by overcoming the limitations of individual barcodes due to PCR biases, low interspecific variability, or the absence of the corresponding sequences in the database. Seethapathy et al. (2019) demonstrated 67% fidelity for targeted plant species present in single drugs of the European market using *ITS1* and *ITS2* barcodes. We obtained 100% fidelity for targeted plant species within single drugs. The overall results of the single drugs revealed that a multi-barcode metabarcoding approach could be used to assess the prevalence of widespread adulterated and substituted plant material in single drugs and implement more stringent supply chain precautionary measures at primary level.

10.3389/fpls.2023.1169984

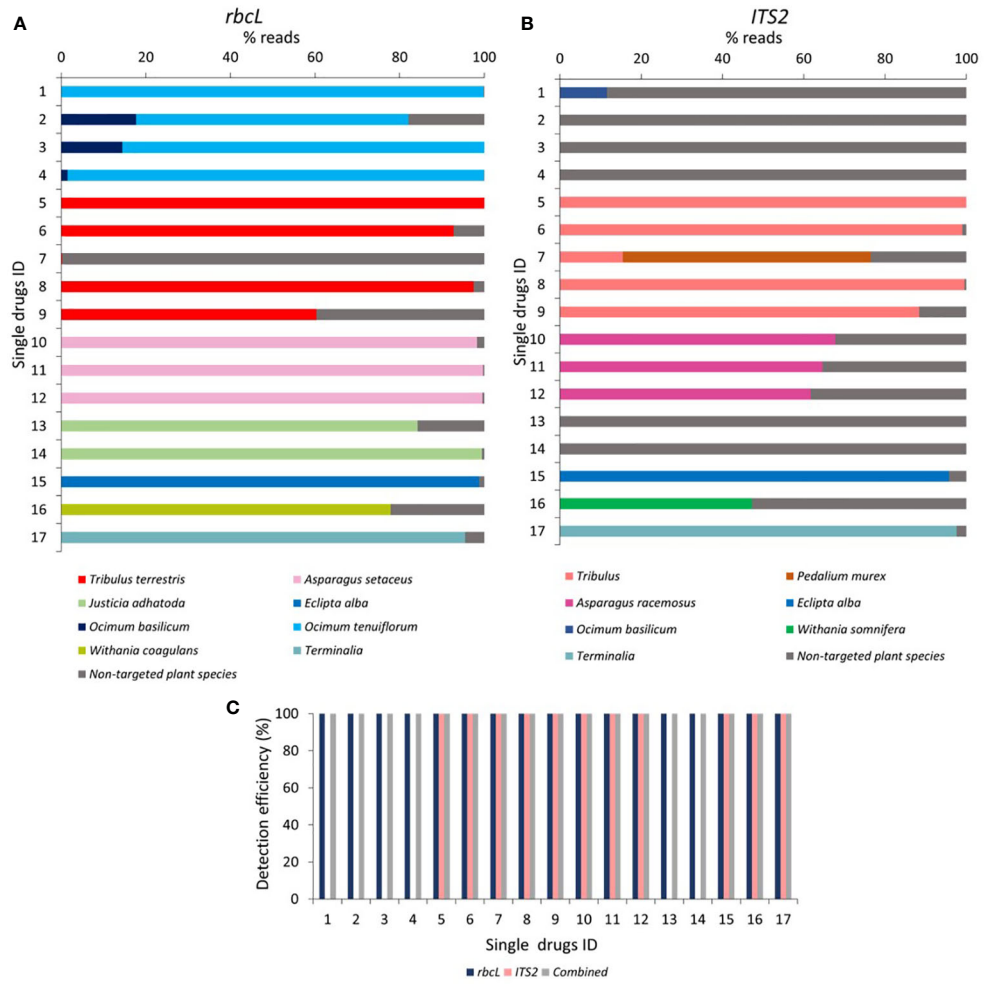


FIGURE 4
Relative abundance of the plant species and detection efficiency in single drugs through *rbcL* and *ITS2* metabarcoding. **(A)** Relative abundance (% reads) of the plant species detected in single drugs through *rbcL* metabarcoding. Relative abundance (% reads) of non-targeted plant species is reported in Supplementary Information Table S4. **(B)** Relative abundance (% reads) of the plant species detected in single drugs through *ITS2* metabarcoding. The relative abundance of non-targeted plant species is reported in Supplementary Information Table S4. **(C)** Detection efficiency obtained in single drugs by *rbcL*, *ITS2*, and combined metabarcoding approach. Single drugs ID 1 to 4 for Tulsi (*Ocimum tenuiflorum*) powder, 5 to 9 for Gokhru (*Tribulus terrestris*) powder, 10 to 12 for Shatavari (*Asparagus racemosus*) powder, 13 and 14 for Vasa (*Justicia adhatoda*) powder, 15 for Bhringraj (*Eclipta alba*) powder, 16 for Ashwagandha (*Withania somnifera*) powder, and 17 for Arjuna (*Terminalia arjuna*) powder.

3.5 Metabarcoding of polyherbal formulations

Trikatu, Sitopaladi, Rasayana, Hingwashtak, and Talisadi (Talisadya) (Joshi et al., 2017) were among the 15 polyherbal formulations collected. Total reads obtained after filtering and percentage of analyzed read (from filtered reads) after discarding OTU clusters having <5 reads for each polyherbal formulation are shown in Supplementary Information Table S5. A total of 53087 and 1429238 reads were obtained by *rbcL* and *ITS2* metabarcode for polyherbal formulation, respectively (Supplementary Information Table S5). On average, 1.4% (0-3.9%) reads have been obtained for non-targeted plant species with *rbcL* metabarcode, and 16.5% (0.1-87.4%) reads were obtained for non-targeted plant species with *ITS2* metabarcode in polyherbal formulations (Supplementary Information Table S6).

In Trikatu, 28.2%, 4.6%, and 57% read for *P. nigrum* with *rbcL*, and 8.3%, 14.7 and 37.9% reads with *ITS2* was observed in sample 18, 19, and 20, respectively. *P.longum* comprised 2%, 37.1%, and 47.8% reads with *rbcL* and 0%, 42.1%, and 19.7% with *ITS2*; *Z. officinale* possess 67.6%, 56.5%, and 43.6% with *rbcL* and 4.3%, 18.8% and 2.4% with *ITS2* in sample 18, 19, and 20 respectively (Figures 5A, B). All three targeted plants were detected (i.e., 100% detection efficiency) in all three Trikatu samples (i.e., 100% fidelity) using a combined approach (Figure 5C, Table 3B). Nevertheless, *ITS2* showed the higher percentage of non-targeted reads (87.4% in sample 18, 24.4% in sample 19, and 40% in sample 20) might be due to technical bias that can be introduced during DNA extraction and PCR towards the unintentional cross-contamination happens during the supply chain (Figure 5B).

Sitopaladi powder is primarily composed of five constituents; with the exclusion of sacrarium (sugar) and vanshlochan (the female bamboo

10.3389/fpls.2023.1169984

TABLE 3 Detection fidelity of single drugs and polyherbal formulations.

TABLE 3A Detection fidelity of single drugs.

Name of Herbal product	Composition of formulations	No. of products	Absolute fidelity[#] /single drug			Relative fidelity (%)/single drug		
			rbcL	*ITS2*	Combined	*rbcL*	*ITS2*	Combined
Tulsi	*Ocimum tenuiflorum*	4	4	0	4	100	0	100
Gokhru	*Tribulus terrestris*	5	5	5	5	100	100	100
Shatavari	*Asparagus racemosus*	3	3	3	3	100	100	100
Vasa	*Justicia adhatoda*	2	2	0	2	100	0	100
Cumulative fidelity of single drugs (%)						**100**	**50**	**100**

[#]Fidelity was not calculated for Bhurgraj, Ashawgandha and Arjuna where n=1.

TABLE 3B Detection fidelity of polyherbal formulations.

Name of Herbal product	Composition of formulations	No. of products	Absolute fidelity of detection/species/ polyherbal formulation[≠]			Relative fidelity of detection/ species/ polyherbal formulation			Average relative fidelity (%)/ polyherbal formulations		
			rbcL	*ITS2*	Combined	*rbcL*	*ITS2*	Combined	*rbcL*	*ITS2*	Combined
Trikatu	*Zinger officinale*	3	3	3	3	100	100	100	88.9	100	100
	Piper nigrum		3	3	3	100	100	100			
	Piper longum		2	3	3	66.7	100	100			
Sitopaladi	*Piper logum*	3	3	3	3	100	100	100	77.8	66.7	77.8
	Eletaria cardamom		3	3	3	100	100	100			
	Cinnamomum cassia		1	0	1	33.3	0	33.3			
Rasayana	*Tribulus terrestris*	4	4	4	4	100	100	100	75	33.3	75
	Tinospora sinensis		4	0	4	100	0	100			
	Phyllanthus emblica		1	0	1	25	0	25			
Hingwashtak	*Zinger officinale*	4	4	1	4	100	25	100	57.1	39.3	63.3
	Piper nigrum		2	0	2	50	0	50			
	Piper longum		1	2	2	25	50	50			
	Apium graveolens (substituted with *Trachyspermum ammi*)		4	3	4	100	75	100			
	Cyminum cyminum		4	4	4	100	100	100			
	Carum carvi [substituted with *Bunium persicum (Elwendia persica)]*		0	1	1	0	25	25			
	Ferula foetida		1	0	1	25	0	25			

[≠]Fidelity was not calculated for Talisadi/Talisadya as n=1.

exudate) and from the total number of designated species, the aim was to detect *C. cassia*, *P. longum*, and *Elettaria cardamom*. *C. cassia* was detected by *rbcL* metabarcode with 0.2% reads only in sample 22. P. *longum* exhibited 1.8, 66.6, and 83.3% of reads with *rbcL* and 0.1, 6.6, and 31.9% of reads with *ITS2* in samples 21, 22, and 23, respectively. *E. cardamomum* showed 96.3%, 30.2%, and 15.9% reads with *rbcL* and 72.4%, 43.7%, and 65.1% with *ITS2* in samples 21, 22, and 23, respectively (Figures 5A, B). Overall, the combined approach showed 66.7% detection efficiency in sample 21, 100% in sample 22, and 66.7% in sample 23, and average of 77.8% fidelity (Figure 5C, Table 3B).

In Rasayana, *rbcL* metabarcode exhibited 91% to 98% reads for *T. teresteris*, while *ITS2* metabarcode exhibited 98 to 100%

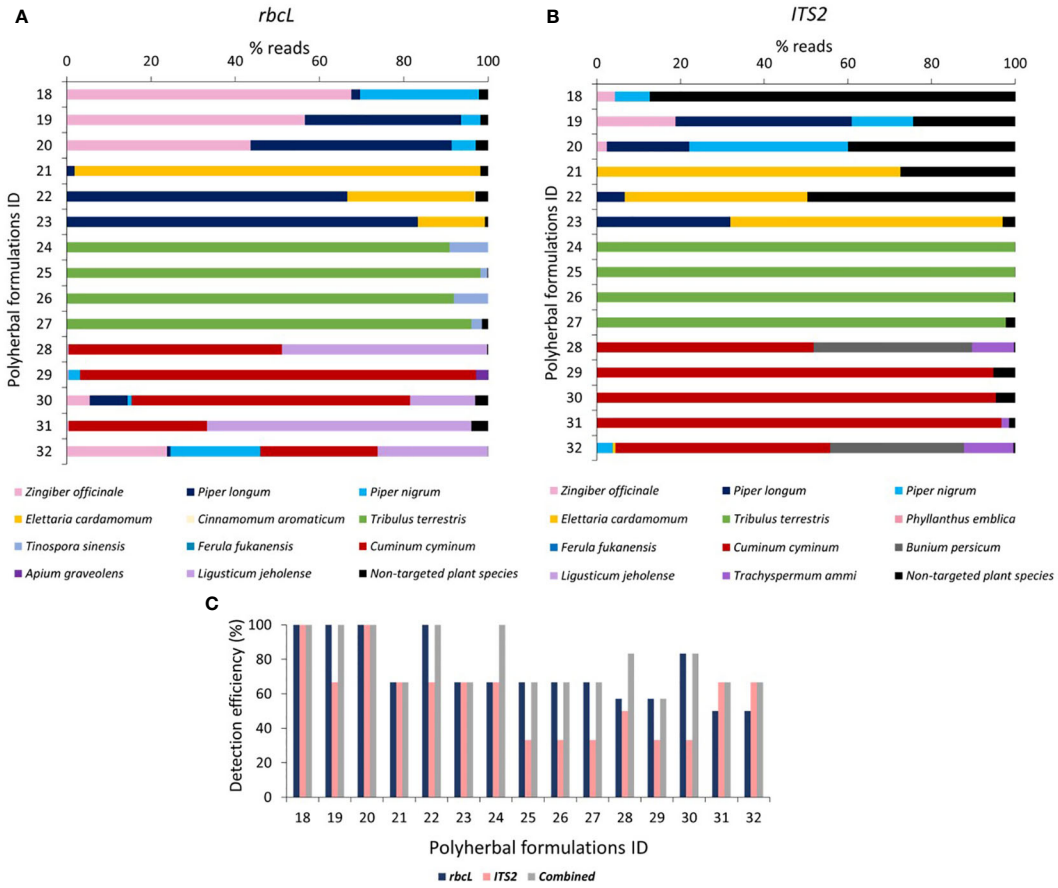


FIGURE 5
Relative abundance of the plant species and detection efficiency in polyherbal formulations through *rbcL* and *ITS2* metabarcoding. **(A)** Relative abundance (% reads) of the plant species detected in polyherbal formulations through *rbcL* metabarcoding. Relative abundance (% reads) of non-targeted plant species is reported in Supplementary Information Table S6. **(B)** Relative abundance (% reads) of the plant species detected in polyherbal formulations through *ITS2* metabarcoding. Relative abundance (% reads) of non-targeted plant species is reported in Supplementary Information Table S6. **(C)** Detection efficiency obtained in polyherbal formulation by *rbcL*, *ITS2*, and combined metabarcoding approach. Polyherbal formulations ID 18 to 20 for Trikatu powder, 21 to 23 for Sitopaladi powder, 24 to 27 for Rasayana powder, 28 to 31 for Hingwashtak powder, and 32 for Talisadi powder.

reads in all samples (samples 24 to 27). *In all samples, T. cordifolia was resolved by* rbcL with 1.7% to 9.1% reads, while *ITS2* metabarcode could not (Figures 5A, B). *P. emblica* was not detected by both metbarcode except in sample 24 (ITS2 *obtained 0.01% reads were)*, possibly due to DNA extraction biases as DNA extraction from amla fruits is difficult due to high acidic nature and high tannin content (Warude et al., 2003). The combined metabarcoding approach showed 100% detection efficiency in sample 24 and 66.7% in remaining other samples with average of 75% fidelity (Figure 5C, Table 3B).

Hingwashtak powder (samples 28 to 31) comprised seven ingredients, including *Zinger officinale, P. nigrum, P. longum, C. cyminum, C. carvi* [*C. cyminum, and C. carvi* commonly substituted with *Bunium persicum* (Syn. *Elwendia persica*) (Johri, 2011; Singh et al., 2017; Bansal et al., 2018)], *Apium leptophyllum* (majority of commercial products comprised/ labeled *Apium graveolens* instead of *A. leptophyllum*; further these plant species commonly substituted with *T. ammi* (Pushpendra et al., 2016)) and *Ferula foetida.* From these seven ingredients, *rbcL* metabarcode was able to resolve *Z. officinale* in all samples with 0.37 to 5.5% reads, *P. nigrum* in sample 29 (2.7% reads) and 30 (1.0% reads), *P. longum* in sample 30 (8.9% reads), *C. cyminum* in all samples with 32.7% to 94.1% reads, *A. graveolens* in sample 29 (2.9% reads). *C. carvi* commonly substituted with *B. persicum* (*Elwendia persica*), neither *C. carvi* nor *B. persicum* was detected by *rbcL* in all Hingwashtak samples (Figure 5A). *ITS2* metabarcode exhibited a high prevalence of *C. cyminum* in all samples with 51.8% to 95.4% reads, *B. persicum* in sample 28 with 37.9% reads, *T. ammi* (substitution of *A. leptophyllum* or *A. graveolens*) in sample 28 with 10% reads (Figure 5B). In addition, *ITS2* metabarcode also detected Trachyspermum ammi in samples 29 and 31. *rbcL* metabarcode showed reads for *Ligusticum jeholense* (Chinese medicinal herb from the Apiaceae family) instead of *Trachyspermum ammi* in sample 28, 30, 31 and *A. graveolens* in sample 29. This could be due to our *rbcL* metabarcode unable to resolve *T. ammi* and detect *L. jeholense* falling under the same family. Overall, the combined metabarcoding approach showed average 72.6%

TABLE 4 Fidelity of targeted plant species present within mock controls as well herbal formulations.

Plant species	Family	Plant part used in simulated biomass mock control or formulations	Resolution at taxa level		A number of gDNA mock controls in which plant species present	Relative fidelity of each plant species presents within first and third type of mock controls (gDNA controls)			A total number of biomass controls, single drugs and polyherbal formulations in which plant species present	Relative fidelity of each plant species presents within second type of mock control (simulated biomass control), single drugs and polyherbal formulations (biomass controls + herbal formulations)			A number of mock controls and herbal products in which plant species present	Relative fidelity of each plant species presents in different types of mock controls and herbal products (cumulative analysis)		
			rbcL	*ITS2*		*rbcL*	*ITS2*	Combined		*rbcL*	*ITS2*	Combined		*rbcL*	*ITS2*	combined
Andrographis paniculata	Acanthaceae	Whole plant	Species	ND	3	100	0	100	3	100.0	0.0	100.0	6	100.0	0.0	100.0
Azadirachcta indica	Meliaceae	Leaves	Species	ND	3	100	33.3	100	3	100.0	66.7	100.0	6	100.0	50.0	100.0
Eclipta alba	Asteraceae	Whole plant	Species	Species	3	100	100	100	4	100.0	100.0	100.0	7	100.0	100.0	100.0
Piper nigrum	Piperaceae	Seed	Species	Species	4	100	100	100	11	81.8	90.9	90.9	15	86.7	93.3	93.3
Zingiber officinale	Zingiberaceae	Rhizome	Family	Family	3	100	100	100	11	54.5	63.6	63.6	14	64.3	71.4	71.4
Aegale marmelous	Rutaceae	Fruits	Family	ND	2	100	50	100	2	50.0	0.0	50.0	4	75.0	25.0	75.0
Centella asiatica	Apiaceae	Whole plant	Species	Species	2	100	100	100	2	100.0	100.0	100.0	4	100.0	100.0	100.0
Terminalia arjuna	Combretaceae	Bark	Genus	Species	3	0	100	100	3	33.3	33.3	33.3	6	16.7	66.7	66.7
Terminalia bellerica	Combretaceae	Fruits	Species	Species	3	100	100	100	2	0.0	100.0	100.0	5	60.0	100.0	100.0
Vitex negundo	Lamiaceae	Leaves	Species	species	2	100	100	100	2	50.0	50.0	50.0	4	75.0	75.0	75.0
Bacopa monneri	Plantaginaceae	Whole plant	Species	Species	1	NA	NA	NA	1	NA	NA	NA	2	50.0	100.0	100.0
Cassia tora	Fabaceae	Leaves	Species	Species	1	NA	NA	NA	1	NA	NA	NA	2	50.0	100.0	100.0
Phyllanthus embilica	Phyllanthaceae	Fruits	Species	Species	2	100	100	100	5	0.0	20.0	20.0	7	28.6	42.9	42.9
Terminalia chebula	Combretaceae	Fruits	Genus	Species	2	100	100	100	1	NA	NA	NA	3	66.7	66.7	100.0
Justicia adhatoda	Acanthaceae	Leaves	Species	ND	1	NA	NA	NA	3	100.0	0.0	100.0	4	100.0	0.0	100.0
Piper longum	Piperaceae	Fruits	Species	Species	1	NA	NA	NA	11	63.6	72.7	72.7	12	58.3	75.0	75.0
Asparagua racemosus	Asparagaceae	Root	Genus	Species	1	NA	NA	NA	3	100.0	100.0	100.0	4	100.0	100.0	100.0

(Continued)

10.3389/fpls.2023.1169984

TABLE 4 Continued

Plant species	Family	Plant part used in simulated biomass mock control or formulations	Resolution at taxa level		A number of gDNA mock controls in which plant species present	Relative fidelity of each plant species presents within first and third type of mock controls (gDNA controls)			A total number of biomass controls, single drugs and polyherbal formulations in which plant species present	Relative fidelity of each plant species presents within second type of mock control (simulated biomass control), single drugs and polyherbal formulations (biomass controls + herbal formulations)			A number of mock controls and herbal products in which plant species present	Relative fidelity of each plant species presents in different types of mock controls and herbal products (cumulative analysis)		
			rbcL	*ITS2*		*rbcL*	*ITS2*	Combined		*rbcL*	*ITS2*	Combined		*rbcL*	*ITS2*	combined
Tribulus terrestris/ Pedalium murex	Zygophyllaceae/Pedaliaceae	Fruits	Species/ ND	Genus/ Species					9	100.0	100.0	100.0	9	100.0	100.0	100.0
Ocimum tenuiflorum/ Ocimum basilicum	Lamiaceae	Leaves	Species/ Species	ND/ ND					4	100.0	0.0	100.0	4	100.0	0.0	100.0
Elettaria cardamomum	Zingiberaceae	Seed	Species	Species					4	75.0	100.0	100.0	4	75.0	100.0	100.0
Carum carvi/ Elwendia persica	Apiaceae	Seed	ND/ ND	ND/ Genus					4	0.0	25.0	25.0	4	0.0	25.0	25.0
Cinnamomum cassia	Lauraceae	Bark	Genus	ND					4	25.0	0.0	25.0	4	25.0	0.0	25.0
Cyminum cyminum	Apiaceae	Seed	Species	Species					4	100.0	100.0	100.0	4	100.0	100.0	100.0
Ferula foetida	Apiaceae	Gum resin	ND	Genus					4	0.0	25.0	25.0	4	0.0	25.0	25.0
Tinospora sinensis	Menispermaceae	Root	Species	ND					4	100.0	0.0	100.0	4	100.0	0.0	100.0
Apium leptophyllum /Trachyspermum ammi/Apium graveolens	Apiaceae	Seed	ND/ ND/ Species	ND/ Species/ ND					4	100.0	75.0	100.0	4	100.0	75.0	100.0
Withania somnifera	Solanaceae	Root	Genus	Species					1	NA	NA	NA	1	NA	NA	NA
Abies webbiana	Pinaceae	Leaves	ND	ND					1	NA	NA	NA	1	NA	NA	NA
Asparagua dumosus	Asparagaceae	Leaves	ND	ND	1	NA	NA	NA					1	NA	NA	NA

(Continued)

10.3389/fpls.2023.1169984

10.3389/fpls.2023.1169984

TABLE 4 Continued

Plant species	Family	Plant part used in simulated biomass mock control or formulations	Resolution at taxa level		A number of gDNA mock controls in which plant species present	Relative fidelity of each plant species presents within first and third type of mock controls (gDNA controls)			A total number of biomass controls, single drugs and polyherbal formulations in which plant species present	Relative fidelity of each plant species presents within second type of mock control (simulated biomass control), single drugs and polyherbal formulations (biomass controls + herbal formulations)			A number of mock controls and herbal products in which plant species present	Relative fidelity of each plant species presents in different types of mock controls and herbal products (cumulative analysis)		
			rbcL	*ITS2*		*rbcL*	*ITS2*	Combined		*rbcL*	*ITS2*	Combined		*rbcL*	*ITS2*	combined
Asparagus adscendens	Asparagaceae	Leaves	ND	Species	1	NA	NA	NA					1	NA	NA	NA
Phyllanthus amaras	Phyllanthaceae	Leaves	ND	Species	1	NA	NA	NA					1	NA	NA	NA
Phyllanthus madresapatensis	Phyllanthaceae	Leaves	ND	ND	1	NA	NA	NA					1	NA	NA	NA
Piper betle	Piperaceae	Leaves	ND	ND	1	NA	NA	NA					1	NA	NA	NA
Pipper arabarum	Piperaceae	Leaves	ND	ND	1	NA	NA	NA					1	NA	NA	NA

ND: Not detected, NA: Species not applicable for fidelity calculation as present only either in one mock control or either in one herbal formulations (n=1). Plant species represent in red colour are substituted plant species.

detection efficiency with 63.3% fidelity for Hingwashtak powders (Figure 5C and Table 3B).

Talisadi/Talisadya powder (sample 32) comprises eight constituents including *Abies webbiana, P. longum, P. nigrum, Z. officinale, E. cardamomum, Cinnamomum Zeylanicum*, Vanshlochan (the female bamboo exudate), and sugar (sugar and Vanshlochan are excluded for metabarcoding analysis). From these six ingredients, only three plant species which include *P. longum* (0.8% reads), *P. nigrum* (21.3% reads), and *Z. officinale* (23.8% reads) were detected by *rbcL* metabarcode. Three plant species which include *P. nigrum* (3.79% reads), *Z. officinale* (0.03% reads), and *E. cardamomum* (0.6% reads) were detected by *ITS2* metabarcode. *A. webbiana* and *C. Zeylanicum* were not detected by either of the barcodes. The combined metabarcode approach detected four plant species (66.7% detection efficiency) out of six (Figure 5C). In Talisadi/Talisadya powder, *C. cyminum* (27.9% reads with *rbcL* and 51.3% reads with *ITS2*), *B. persicum* (32.0% reads with *ITS2*), *L. jeholense* (26.2% reads with *rbcL*) and *T. ammi* (11.8% reads with *ITS2*) was detected might be due to unintentional cross-contamination happen during sample processing as a collection of Hingwashtak powder sample 28 and Talisadi powder sample 32 were done from the same company. In addition, a high percentage of reads were covered by *C. cyminum*, *T. ammi*, *L. jeholens*, *B. persicum* (all plant species belonging to Apiaceae family), then *Z. officinale*, *P. longum*, and *P. nigrum*. That could be because of technical bias introduced during DNA extraction and PCR.

3.6 Fidelity of targeted plant species

Up to this point, the fidelity of plant species per number of mock controls or herbal formulations has been calculated and discussed. Here, we have estimated the fidelity of the targeted plant species included within different mock controls and polyherbal formulations to get a better perspective of species discrimination capabilities and reliabilities of single and multi-barcode approaches. Both the *rbcL* and *ITS2* metabarcodes resolved 19 (46.7%) of the 39 listed plant species at the species level. However, plants detected at the species level were different, and a multi-barcode approach provided species-level resolution for 27 (69.23%) species, leading to a 20.5% increment in the whole (Table 4). This observation confirms robustness for our newly designed metabarcodes in detecting plant species at lower taxonomic levels. In addition, 100% fidelity was observed for *T. bellirica, A. paniculata, A. indica, E. alba*, and *C. asiatica* within gDNA controls, biomass controls, and cumulative analysis by the combined approach of *rbcL* and *ITS2*. However, the combined approach of *rbcL* and *ITS2* exhibited 100% fidelity for *Z. officinale*, *V. negundo*, *P. nigrum*, *A. marmelous*, *T. arjuna*, and *P. embilica* only within gDNA controls and having lower fidelity in biomass controls and cumulative analysis (Table 4). That could be due to biases in the DNA isolation process; yielding equal proportional DNA from the poly formulation is impossible due to genome size differences and differences in plant parts and amounts of secondary metabolites.

Furthermore, the extracted DNA is degraded because herbal products are intensively processed. The PCR conditions and reactions will also have a significant impact on the primer fit and PCR bias of the mixture. That was demonstrated by comparing the combined fidelity of gDNA controls, biomass controls, and cumulative analysis (Table 4). On average, 83.6% fidelity was observed for targeted plant species in the cumulative analysis. This result confirmed the high reliability of our multi-barcode sequencing approach.

4 Conclusion

On the whole, our findings suggest that the multi-barcode DNA metabarcoding method assessed in this study can provide a composition of more diverse sets of single drugs and polyherbal formulations listed in the Ayurvedic Pharmacopoeia of India. We obtained 100% average detection efficiency and relative fidelity of targeted plants for single drugs and 79% for polyherbal formulations through the multi-barcode sequencing approach. We have primarily focused on detected plant species in herbal products rather than undetected plant species because many steps, such as DNA extraction biases, PCR biases, and manufacturing processes, that can lead to DNA degradation or loss beyond detectable limits, failing to detect plant species. The presence of non-targeted plant species in herbal products could be due to unintentional contamination of the supply chain, economically motivated adulteration, and/or admixture of other species. Our study showed that the *rbcL* metabarcode had better detection ability for certain plant species, e.g., *O. tenuiflorum, J. adhatoda*, and *A. paniculata*, while *ITS2* had better discrimination power for certain plant species, e.g., species of the genus *Terminalia*, *Asparagus, Piper*, *Phyllanthus*, and *Pedalium murex*. Thus, the complementary approach of both metabarcodes is a promising tool for quality evaluation of herbal products and pharmacovigilance. However, the development of standardized methods for metabarcoding sequencing and bioinformatics analysis pipeline and curated database is needed for effective use as a regulatory tool to authenticate herbal products in combination with advanced chemical methods to identify bioactive therapeutics.

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.

Author contributions

TT: performed experiments, established metabarcoding data analysis pipeline, data analysis, writing and editing manuscript, and validation of final manuscript. AS: performed experiments, data analysis, writing and editing manuscript, and validation of final manuscript. RP: designed primers for metabarcoding, performed experiments, established metabarcoding data analysis pipeline, and manuscript editing. SS: performed experiments, and manuscript editing. CJ: project administration, methodology, supervision, and review & editing. MJ: principal investigator, conceptualization, methodology, supervision, and review & editing. All authors contributed to the article and approved the submitted version.

Funding

Gujarat State Biotechnology Mission (GSBTM), Gandhinagar, Gujarat, India, has provided financial support for the project under the Research Support Scheme, grant ID GSBTM/JDRD/584/2018/204.

Acknowledgments

The authors would like to thank Prof. Padamnabhi S. Nagar, The Maharaja Sayajirao University of Baroda, Gujarat, India and Director, DMAPR, Gujarat, India for helping us in plant collection and authentication. The authors would like to thank Dr. Darshan Parmar, M.D. (Rasashastra and Bhaishajya Kalpana), Government Ayurvedic College, Vadodara (Gujarat, India) for providing some of the herbal formulations and Mr. Nitin Savaliya, Technical Assistant from Thermo Fisher Scientific, for NGS instrument handling and run setup.

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1169984/full#supplementary-material

References

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2

Arulandhu, A. J., Staats, M., Hagelaar, R., Voorhuijzen, M. M., Prins, T. W., Scholtens, I., et al. (2017). Development and validation of a multi-locus DNA metabarcoding method to identify endangered species in complex samples. *Gigascience* 6 (10), gix080. doi: 10.1093/gigascience/gix080

Bansal, S., Thakur, S., Mangal, M., Mangal, A. K., and Gupta, R. K. (2018). DNA Barcoding for specific and sensitive detection of cuminum cyminum adulteration in bunium persicum. *Phytomedicine* 50, 178–183. doi: 10.1016/j.phymed.2018.04.023

Cheng, X., Su, X., Chen, X., Zhao, H., Bo, C., Xu, J., et al. (2014). Biological ingredient analysis of traditional Chinese medicine preparation based on high-throughput sequencing: the story for liuwei dihuang wan. *Sci. Rep.* 4, 1–12. doi: 10.1038/srep05147

Choudhary, S., Kaurav, H., and Chaudhary, G. (2021). Gokhru (tribulus terrestris and pedalium murex): medicinal importance of chota gokhru and bada gokhru in ayurveda and modern science. *Asian J. Pharm. Clin. Res.* 14, 6–13. doi: 10.22159/ajpcr.2021.v14i6.41366

Coghlan, M. L., Haile, J., Houston, J., Murray, D. C., White, N. E., Moolhuijzen, P., et al. (2012). Deep sequencing of plant and animal DNA contained within traditional Chinese medicines reveals legality issues and health safety concerns. *PloS Genet.* 8 (4), e1002657. doi: 10.1371/journal.pgen.1002657

Fazekas, A. J., Kesanakurti, P. R., Burgess, K. S., Percy, D. M., Graham, S. W., Barrett, S. C. H., et al. (2009). Are plant species inherently harder to discriminate than animal species using DNA barcoding markers? *Mol. Ecol. Resour.* 9, 130–139. doi: 10.1111/j.1755-0998.2009.02652.x

Frigerio, J., Agostinetto, G., Mezzasalma, V., De Mattia, F., Labra, M., and Bruno, A. (2021). Dna-based herbal 'teas' authentication: an *ITS2* and psba-trnh multi-marker dna metabarcoding approach. *Plants* 10, 1–14. doi: 10.3390/plants10102120

Hebert, P. D. N., Cywinska, A., Ball, S. L., and DeWaard, J. R. (2003). Biological identifications through DNA barcodes. *Proc. R. Soc B Biol. Sci.* 270, 313–321. doi: 10.1098/rspb.2002.2218

Huang, Y., Niu, B., Gao, Y., Fu, L., and Li, W. (2010). CD-HIT suite: a web server for clustering and comparing biological sequences. *Bioinformatics* 26, 680–682. doi: 10.1093/bioinformatics/btq003

Ichim, M. C. (2019). The DNA-based authentication of commercial herbal products reveals their globally widespread adulteration. *Front. Pharmacol.* 10. doi: 10.3389/fphar.2019.01227

Ivanova, N. V., Kuzmina, M. L., Braukmann, T. W. A., Borisenko, A. V., and Zakharov, E. V. (2016). Authentication of herbal supplements using next-generation sequencing. *PloS One* 11, 1–24. doi: 10.1371/journal.pone.0156426

Johri, R. K. (2011). Cuminum cyminum and carum carvi: an update. *Pharmacogn. Rev.* 5, 63–72. doi: 10.4103/0973-7847.79101

Joshi, V. K., Joshi, A., and Dhiman, K. S. (2017). The ayurvedic pharmacopoeia of India, development and perspectives. *J. Ethnopharmacol.* 197, 32–38. doi: 10.1016/j.jep.2016.07.030

Kress, W. J. (2017). Plant DNA barcodes: applications today and in the future. *J. Syst. Evol.* 55, 291–307. doi: 10.1111/jse.12254

Linhart, C., and Shamir, R. (2005). The degenerate primer design problem: theory and applications. *J. Comput. Biol.* 12, 431–456. doi: 10.1089/cmb.2005.12.431

Lo, Y. T., and Shaw, P. C. (2019). Application of next-generation sequencing for the identification of herbal products. *Biotechnol. Adv.* 37, 107450. doi: 10.1016/j.biotechadv.2019.107450

Maloukh, L., Kumarappan, A., Jarrar, M., Salehi, J., El-wakil, H., and Rajya Lakshmi, T. V. (2017). Discriminatory power of *rbcL* barcode locus for authentication of some of united Arab Emirates (UAE) native plants. *3 Biotech.* 7, 1–7. doi: 10.1007/s13205-017-0746-1

Mishra, P., Kumar, A., Nagireddy, A., Mani, D. N., Shukla, A. K., Tiwari, R., et al. (2016). DNA Barcoding: an efficient tool to overcome authentication challenges in the herbal market. *Plant Biotechnol. J.* 14, 8–21. doi: 10.1111/pbi.12419

Newmaster, S. G., Grguric, M., Shanmughanandhan, D., Ramalingam, S., and Ragupathy, S. (2013). DNA Barcoding detects contamination and substitution in north American herbal products. *BMC Med.* 11, 222. doi: 10.1186/1741-7015-11-222

Pandit, R., Travadi, T., Sharma, S., Joshi, C., and Joshi, M. (2021). DNA Meta-barcoding using *rbcL* based mini-barcode revealed presence of unspecified plant species in ayurvedic polyherbal formulations. *Phytochem. Anal.* 32, 804–810. doi: 10.1002/pca.3026

Parveen, I., Gafner, S., Techen, N., Murch, S. J., and Khan, I. A. (2016). DNA Barcoding for the identification of botanicals in herbal medicine and dietary supplements: strengths and limitations. *Planta Med.* 82, 1225–1235. doi: 10.1055/s-0042-111208

Pushpendra, P., Sunil Kumar, K. N., Priyadarshini, P., Holla, B. S., Ravishankar, B., and Yashovarma, B. (2016). Quality standards for hutabhuga¯di cu¯ra (Ayurvedic formulary of India). *J. Tradit. Complement. Med.* 6, 78–88. doi: 10.1016/j.jtcme.2014.11.019

Raclariu, A. C., Heinrich, M., Ichim, M. C., and de Boer, H. (2018a). Benefits and limitations of DNA barcoding and metabarcoding in herbal product authentication. *Phytochem. Anal.* 29, 123–128. doi: 10.1002/pca.2732

Raclariu, A. C., Mocan, A., Popa, M. O., Vlase, L., Ichim, M. C., Crisan, G., et al. (2017a). Veronica officinalis product authentication using DNA metabarcoding and HPLC-MS reveals widespread adulteration with veronica chamaedrys. *Front. Pharmacol.* 8, 378. doi: 10.3389/fphar.2017.00378

Raclariu, A. C., Paltinean, R., Vlase, L., Labarre, A., Manzanilla, V., Ichim, M. C., et al. (2017b). Comparative authentication of hypericum perforatum herbal products using DNA metabarcoding, TLC and HPLC-MS. *Sci. Rep.* 7, 8–10. doi: 10.1038/s41598-017-01389-w

Raclariu, A. C., Ţebrencu, C. E., Ichim, M. C., Ciupercă, O. T., Brysting, A. K., and de Boer, H. (2018b). 'What's in the box? authentication of echinacea herbal products using DNA metabarcoding and HPTLC. *Phytomedicine* 44, 32–38. doi: 10.1016/j.phymed.2018.03.058

Schmieder, R., and Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 27, 863–864. doi: 10.1093/bioinformatics/btr026

Seethapathy, G. S., Raclariu-Manolica, A.-C., Anmarkrud, J. A., Wangensteen, H., and De Boer, H. J. (2019). DNA Metabarcoding authentication of ayurvedic herbal products on the European market raises concerns of quality and fidelity. *Front. Plant Sci.* 10, 68. doi: 10.3389/fpls.2019.00068

Sharma, S., and Shrivastava, N. (2016). Internal transcribed spacer guided multiplex PCR for species identification of convolvulus prostratus and evolvulus alsinoides. *Acta Pharm. Sin. B* 6, 253–258. doi: 10.1016/j.apsb.2016.02.003

Shetti, S., Kumar, C. D., Sriwastava, N. K., and Sharma, I. P. (2011). Pharmacovigilance of herbal medicines: current state and future directions. *Pharmacogn. Mag.* 7, 69–73. doi: 10.4103/0973-1296.75905

Singh, R. P., H.V., G., and K, M. (2017). *Cuminum cyminum* {\textendash} a popular spice: an updated review. *Pharmacogn. J.* 9, 292–301. doi: 10.5530/pj.2017.3.51

The ayurvedic pharmacopoeia of India, 1s, Vol. I. (1990). Department of AYUSH, Ministry of Health and Family Welfare, Government of India.

The ayurvedic pharmacopoeia of India, 1st, Vol. II. (1999). Department of AYUSH, Ministry of Health and Family Welfare, Government of India.

The ayurvedic pharmacopoeia of India, 1st, Vol. III. (2001). Department of AYUSH, Ministry of Health and Family Welfare, Government of India.

Travadi, T., Shah, A. P., Pandit, R., Sharma, S., Joshi, C., and Joshi, M. (2022a). Detection of carica papaya adulteration in piper nigrum using chloroplast DNA marker-based PCR assays. *Food Anal. Methods.* 16, 107–114. doi: 10.1007/s12161-022-02395-z

Travadi, T., Sharma, S., Pandit, R., Nakrani, M., Joshi, C., and Joshi, M. (2022b). A duplex PCR assay for authentication of ocimum basilicum l. and ocimum tenuiflorum l in tulsi churna. *Food Control* 137, 108790. doi: 10.1016/j.foodcont.2021.108790

Urumarudappa, S. K. J., Tungphatthong, C., Prombutara, P., and Sukrong, S. (2020). DNA Metabarcoding to unravel plant species composition in selected herbal medicines on the national list of essential medicines (NLEM) of Thailand. *Sci. Rep.* 10, 1–11. doi: 10.1038/s41598-020-75305-0

Vassou, S. L., Nithaniyal, S., Raju, B., and Parani, M. (2016). Creation of reference DNA barcode library and authentication of medicinal plant raw drugs used in ayurvedic medicine. *BMC Complement. Altern. Med.* 16, 9–15. doi: 10.1186/s12906-016-1086-0

Ved, D. K., and Goraya, G. S. (2007). Demand and supply of medicinal plants in India. NMPB, New Delhi & FRLHT, Bangalore, India. 18 (85), 210–52.

Warude, D., Chavan, P., Joshi, K., and Patwardhan, B. (2003). DNA Isolation from fresh and dry plant samples with highly acidic tissue extracts. *Plant Mol. Biol. Rep.* 21, 467. doi: 10.1007/BF02772600

Wu, H. Y., and Shaw, P. C. (2022). Strategies for molecular authentication of herbal products: from experimental design to data analysis. *Chin. Med. (United Kingdom)* 17, 1–15. doi: 10.1186/s13020-022-00590-y

Xin, T., Xu, Z., Jia, J., Leon, C., Hu, S., Lin, Y., et al. (2018). Biomonitoring for traditional herbal medicinal products using DNA metabarcoding and single molecule, real-time sequencing. *Acta Pharm. Sin. B* 8, 488–497. doi: 10.1016/j.apsb.2017.10.001

Yao, Q., Zhu, X., Han, M., Chen, C., Li, W., Bai, H., et al. (2022). Decoding herbal materials of TCM preparations with the multi-barcode sequencing approach. *Sci. Rep.* 12, 1–18. doi: 10.1038/s41598-022-09979-z

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 31 May 2023
DOI 10.3389/fpls.2023.1178778

OPEN ACCESS

EDITED BY
Yinglong Chen,
University of Western Australia, Australia

REVIEWED BY
Zhihua Hua,
Ohio University, United States
Han Xiao,
Center for Excellence in Molecular Plant Sciences (CAS), China

*CORRESPONDENCE
Bram Van de Poel
bram.vandepoel@kuleuven.be

RECEIVED 03 March 2023
ACCEPTED 04 May 2023
PUBLISHED 31 May 2023

CITATION
Geldhof B, Pattyn J and Van de Poel B (2023) From a different angle: genetic diversity underlies differentiation of waterlogging-induced epinasty in tomato. *Front. Plant Sci.* 14:1178778. doi: 10.3389/fpls.2023.1178778

From a different angle: genetic diversity underlies differentiation of waterlogging-induced epinasty in tomato

Batist Geldhof[1], Jolien Pattyn[1] and Bram Van de Poel[1,2]*

[1]Molecular Plant Hormone Physiology Lab, Division of Crop Biotechnics, Department of Biosystems, KU Leuven, Leuven, Belgium, [2]KU Leuven Plant Institute (LPI), KU Leuven, Leuven, Belgium

In tomato, downward leaf bending is a morphological adaptation towards waterlogging, which has been shown to induce a range of metabolic and hormonal changes. This kind of functional trait is often the result of a complex interplay of regulatory processes starting at the gene level, gated through a plethora of signaling cascades and modulated by environmental cues. Through phenotypical screening of a population of 54 tomato accessions in a Genome Wide Association Study (GWAS), we have identified target genes potentially involved in plant growth and survival during waterlogging and subsequent recovery. Changes in both plant growth rate and epinastic descriptors revealed several associations to genes possibly supporting metabolic activity in low oxygen conditions in the root zone. In addition to this general reprogramming, some of the targets were specifically associated to leaf angle dynamics, indicating these genes might play a role in the induction, maintenance or recovery of differential petiole elongation in tomato during waterlogging.

KEYWORDS

waterlogging, tomato, GWAS, epinasty, abiotic stress, ontogeny, real-time data

Introduction

Leaf angle is an important agricultural trait in both monocot and dicot crops that ensures optimal light interception and resource allocation for plant growth and development. Especially for cereal crops, growing in high densities, this feature has gained a lot of attention as a possible target for yield optimization (Mantilla-Perez and Salas Fernandez, 2017; Mantilla-Perez et al., 2020). Leaf angle is, however, a dynamic trait, and plants can adjust their leaf posture to enhance survival in suboptimal conditions. Arabidopsis, for example, wields the hyponastic response as a strategy to overcome low-oxygen stress (Hebelstrup et al., 2012; Rauf et al., 2013), to compete with neighbors by means of the shade-avoidance response (Pantazopoulou et al., 2017) and to improve heat dissipation during high temperatures (Park et al., 2019). This upwards bending is the outcome of an array of hormonal cues that control leaf angle in space and time. Several key

regulators, transducing hormonal signals through genetic programming, have been described, revealing a common framework for posture control during abiotic stress in Arabidopsis.

Many of the known regulators of hyponasty are involved in hormonal signaling. During waterlogging in Arabidopsis, hyponastic growth is triggered by SPEEDY HYPONASTIC GROWTH (SHYG), which acts upstream of 1-AMINOCYCLOPROPANE-1-CARBOXYLIC ACID (ACC) OXIDASE5 (ACO5), a key enzyme in ethylene biosynthesis (Rauf et al., 2013). Ethylene induced hyponasty itself is further controlled by reorientation of cortical microtubules, leading to differential elongation in the petiole (Polko et al., 2012). This process is likely further regulated by brassinosteroid (BR) signaling (Polko et al., 2013). Concomitantly, key photomorphogenic transcription factors are activated to, independent of light signaling, induce upward leaf movement in *Rumex palustris*, indicating shade avoidance and low-oxygen responses rely on a shared arsenal of regulators (Cox et al., 2004; van Veen et al., 2013).

In contrast to Arabidopsis wielding hyponasty as a stress response, other species such as tomato, actively bend their leaves downwards during waterlogging, which is called epinasty. It is likely that the epinastic response is also the output of a series of hormone-regulated processes, starting with the accumulation of ACC in the hypoxic roots (Olson et al., 1995; Shiu et al., 1998). ACC is then transported to the shoot (Jackson and Campbell, 1976; English et al., 1995) and converted into ethylene (English et al., 1995), which converges on AUXIN/INDOLE-3-ACETIC ACID 3 (SlIAA3), an integrator of ethylene and auxin responses that is partially controlling leaf epinasty (Chaabouni et al., 2009). As a result, differential elongation is likely, but not exclusively, gated by auxin redistribution (Lee et al., 2008). *SlIAA3* expression is also modulated by BR, as impaired functioning of DWARF (DWF), a BR biosynthesis gene, represses both ethylene and BR production, strongly induces *SlIAA3* expression (in the stem) and evokes a hyponastic phenotype (Li et al., 2016).

In contrast to the genetic networks resulting in asymmetric petiole growth during hyponasty in Arabidopsis, little is known about the central nodes of such a regulatory network in waterlogging-induced epinasty in tomato. In search for these regulators, we performed a multi-trait and longitudinal (time-series) Genome-Wide Association Study (GWAS) in tomato, revealing that leaf posture control is a highly dynamic trait controlled by many factors. This study shows that tomato harbors genetic diversity with respect to leaf epinasty as an adaptive strategy to survive waterlogging.

Materials and methods

Plant material and growth conditions

Tomato seeds (*Solanum lycopersicum*) of 54 accessions, part of a larger collection described in Tieman et al. (2017), were obtained from the Tomato Genetics Resource Center (TGRC) or were kindly provided by Prof. D. Zamir. Seeds were germinated in soil and afterwards transferred to rockwool blocks. Tomato seedlings were grown in the greenhouse until they reached approximately the eighth leaf stage, after which they were used for experiments. In the greenhouse, temperature was set at 18°C (day and night) with a relative humidity between 65 – 70%. Additional illumination (SON-T) was provided if solar light intensity dropped below 250 W m^{-2}. The plants received fertigation through automated drip irrigation.

Waterlogging treatment and phenotyping

At the eighth leaf stage, half of the plants of each accession was transferred to individual containers and subjected to a three-day waterlogging treatment. The container was filled with water up to 4 cm above the rockwool surface. Oxygen depletion during waterlogging was achieved by natural root respiration, leading to rapid hypoxia, generally within 3 – 8 h (Geldhof et al., 2021). Afterwards, the root system was drained and reoxygenated and plants were monitored for another three days. At the start of the treatment and during reoxygenation, plant height was measured daily. Leaf angle dynamics were monitored using real-time leaf angle sensors (Geldhof et al., 2021), attached to the petiole of the first eight leaves to take into account ontogenic differences. Four out of the 54 accessions were not shared for growth and leaf angle phenotyping (52 unique accessions per dataset).

Genome-wide association study

Time-series continuous leaf angle data were transformed into key traits describing leaf angle dynamics (see results and Supplemental Tables S1, S1B). GWA mapping was performed based on the study by Tieman et al. (2017), using the Efficient Mixed-Model Association eXpedited (EMMAX) algorithm (with missingness < 10% and minor allele frequency (maf) > 5%) (Kang et al., 2010). Monoallelic SNP files were filtered, only retaining the accessions used in this study, and converted into biallelic and binary format using PLINK. The significant P-value threshold was determined using the Genetic type 1 Error Calculator (GEC). The power of the analysis was verified using randomized phenotypes and potential environmental interactions were tested with the Genome-wide Efficient Mixed Model Association (GEMMA) algorithm (Zhou and Stephens, 2014). Design parameters for the model (genotype – phenotype combinations) were bash scripted in R (R Core Team, 2019) and the analysis was run for each of the above phenotypical variables and for the leaf angle itself at 10 minute intervals.

Plant growth during waterlogging and recovery were defined as the average growth over the 3-day waterlogging period or as the slope of a linear model during recovery respectively. The ontogenic effect was incorporated in the GWAS through quadratic models with leaf age as independent and angle descriptors as dependent variables. Leaf age-specific time-series were represented using Functional Principal Component Analysis (FPCA) on downsampled series, angular descriptors and their Best Linear Unbiased Estimators (BLUE). For each of these variables, average values were determined per treatment and their differences were fed to the EMMAX algorithm.

RNA extraction and RT-qPCR

Tomato petioles were sampled and adaxial and abaxial sections were snap frozen and ground in liquid nitrogen (5 replicates per section). RNA was extracted using the GeneJET Plant RNA Purification Mini Kit protocol (Thermo Scientific). DNA was removed using the RapidOut DNA Removal Kit (Thermo Scientific). Total RNA yield was assessed on the Nanodrop (Nanodrop Technology) and RNA quality was verified through gel electrophoresis. cDNA was synthesized using the iScript cDNA Synthesis Kit protocol (BIO-RAD). RT-qPCR was performed with a StepOnePlus (Applied Biosystems) for 40 cycles (SsoAdvanced Universal SYBR Green Supermix; Bio-Rad). Relative quantification and PCR efficiency was calculated using a standard dilution series for each primer-pair. Four reference genes (EF1a, PP2AC, RPL2 and TIP41) were selected based on earlier findings (van de Poel et al., 2012) and their mean expression was used for normalization. Primers are listed in Supplemental Table S2.

Data analysis

All analyses and visualizations were performed in R. Outliers of phenotypical variables were removed prior to the GWAS. FPCA was carried out with the *fdapace* package (Zhou et al., 2022). Gene expression levels were compared between control and waterlogging treatment with a Wilcoxon test ($\alpha = 0.05$). Data visualization was done with the *ggplot* package (Wickham, 2009).

Results

Growth dynamics during waterlogging reveals an accession-dependent pause-recovery strategy

Waterlogging results in suboptimal conditions eventually leading to reduced growth. We quantified this growth effect for 52 tomato accessions after 3 days of waterlogging (Figures 1A, B) and during the subsequent 3-day recovery phase (Figure 1C, Supplemental Figure S1) by measuring changes in plant height. While most of these accessions had a reduced growth rate, some of them retained a more steady growth during waterlogging (e.g. accession 53). Other accessions showed a growth pause during waterlogging, followed by enhanced growth during recovery (e.g. accessions 10, 13, 38, below the identity line in Figure 1C) or seemed unaffected in both conditions (e.g. accessions 21, 43, 47). The extent of this growth reduction was in part determined by the initial plant size, indicating that plant development guides waterlogging responses of tomato (significant effect; Supplemental Figure S2). Overall, the outgrowth of new leaves was delayed for all accessions, but with differences between accessions (Figure 1D). The effect of waterlogging and reoxygenation is clearly visible in the growth dynamics of certain accessions (exemplified in Figure 1E for the accessions below the identity line in Figure 1C).

Growth during waterlogging and recovery is genetically regulated

The phenotypic variability in growth reduction and reinitiation indicates that different accessions wield different strategies to overcome waterlogging. We next investigated if variability of this resilience or recovery potential is associated with single nucleotide polymorphisms (SNPs). Therefore, both differential growth over the entire waterlogging ($G_C - G_H$; C = control, H = hypoxia) and subsequent recovery ($G_C - G_R$; C = control, R = recovery) treatment were used as input phenotypes for a GWAS analysis. As the 52 accessions used in this study originated from a larger population of 402 accessions (described in Tieman et al. (2017)), we decided to incorporate the overarching population genetic variability in the analysis. To this extent, principle component analysis (PCA) scores were derived based on the SNP data for the entire population and the PCs of the 52 accessions were retained as covariates in the analysis. This methodology allowed us to make full use of the underlying population structure (Figure 2A), largely covered by the 52 accessions, except for a subgroup of divergent heirloom varieties.

The results of the GWAS were validated based on the Manhattan and QQplot plot (Figures 2B, C). Despite deflation ($\lambda = 0.907$ & $\lambda = 0.894$), the analysis picked up targets exceeding both the suggestive and significant $-\log_{10}(P)$ values ($P_{sug} = 5.50$ and $P_{sign} = 6.80$), indicating these SNPs could be associated with growth regulation during waterlogging. The SNPs that could be annotated were retained and are listed in (Supplemental Table S3). Two of these SNPs reoccurred during both waterlogging and recovery. The first SNP was located in the 3' UTR region of the gene *BTB/POZ domain containing protein expressed* (Solyc05g052960), a homolog of the *BTB/POZ and MATH domain-containing protein 3* (*BPM3*) gene in Arabidopsis (AT2G39760), known to target HOMEOBOX PROTEIN 6 (AtHB6) to regulate growth and abscisic acid (ABA) signaling (Lechner et al., 2011). The second SNP was located in an exon of a BAHD acyltransferase DCR/HXXXD-type acyl-transferase/hydroxycinnamoyl transferase family protein gene (Solyc05g052680; ortholog of AT2G39980), possibly involved in specialized metabolism.

Waterlogging-induced epinasty as a conserved trait in tomato

The above analysis revealed that growth during waterlogging is genetically determined, leading to different survival strategies. One strategy is to maintain growth during waterlogging and recovery, while another is the induction of a pause phase, probably encompassing morphological adaptations such as the formation of aerenchyma and/or adventitious roots. Another morphological adaptation towards root hypoxia is leaf epinasty. To explore the genotype-specific variability in angular dynamics during waterlogging, we monitored leaf movements with digital sensors (Geldhof et al., 2021), taking into account leaf age. High resolution time-series data were compressed in 14 different traits and their derivative variables (e.g. best linear unbiased estimators (BLUEs))

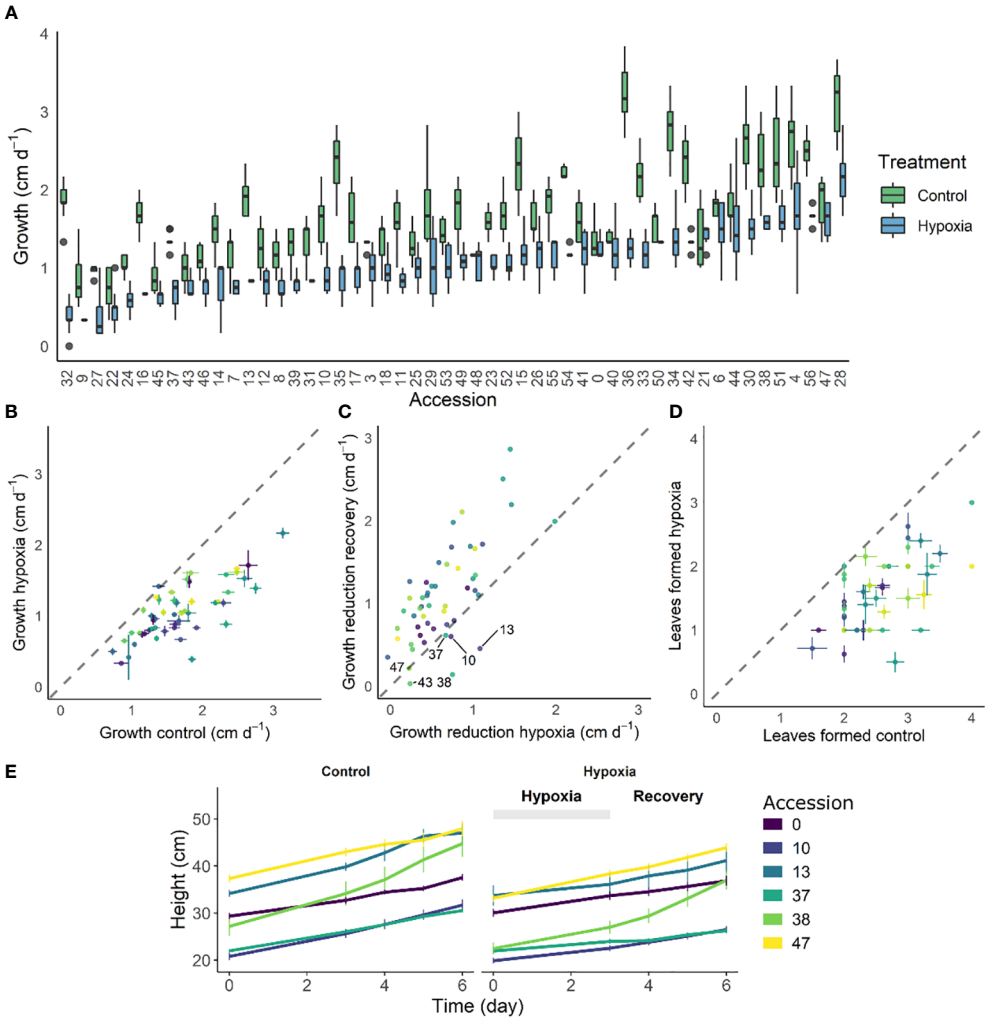


FIGURE 1
Effect of 3 days of waterlogging and recovery on growth of 52 different tomato accessions. **(A)** Natural variation in growth rate under control conditions and during waterlogging (n = 52). **(B, C)** Shift in average growth rate of the accessions during **(B)** 3 days of waterlogging and **(C)** subsequent 3 days of recovery. **(D)** Effect of waterlogging and recovery on leaf emergence for the different accessions. **(E)** Growth curves of 6 accessions during waterlogging and recovery as compared to control conditions. Vertical and horizontal bars in **(A, C, E)** are 95% confidence interval.

describing the epinastic response (Figures 3A, B), including total angular change during hypoxia and the maximal rate of change during the entire sequence (hypoxia and recovery). We first verified the usability of these descriptors by PCA, demonstrating that they could distinguish phenotypic variation during the treatment (Figures 3C, D). On the other hand, many of the traits were grouped, indicating they were inherently correlated. For example, the total angular change ($\theta_{end} - \theta_0$) and rate of change during hypoxia ($\Delta(\theta)_{max}$) were associated, meaning that fast epinastic responses and large epinastic curvatures go hand in hand. Interestingly, this does not necessarily imply that these plants had a strong recovery potential after the treatment.

The above descriptors are not only capable of describing waterlogging-induced changes in leaf posture, but they also vary between different accessions (Figures 3E, F), providing a phenotypical fingerprint for the GWAS. An example of the effect of waterlogging on total leaf angle change and related leaf posture traits for leaf 5 of 52 accessions is given in Figure 3E and Supplemental Figure S3 respectively. While some of the accessions show major leaf repositioning throughout waterlogging and recovery (e.g. 14, 18, 42), others seem to recover (e.g. 10, 30) (Figure 3E). By combining all of these angular variables (Figure 3B), an epinastic heatmap of the waterlogging effect could be devised for each of the accessions (Supplemental Figure S3). This genetic variability, leading to distinctive dynamic leaf movements, is exemplified for different accessions (Figure 3F) and developmental stages (leaf 3, 5 and 7), indicating that both genotype and ontogeny guide leaf movement during waterlogging and recovery.

Angular dynamics during waterlogging is associated with distinct SNP signatures

Similarities in leaf posture and movements between different tomato accessions indicate that epinasty is a conserved response

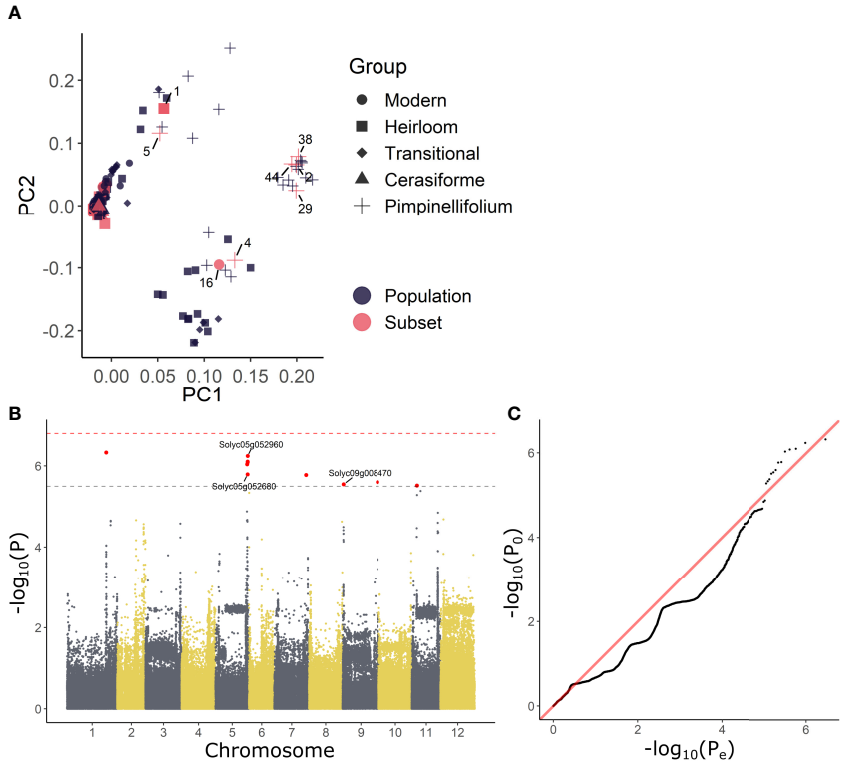


FIGURE 2
GWAS analysis on growth responses during waterlogging of tomato. **(A)** Population structure of the 54 tomato accessions subset used in this study (red) and those of the full population (402 accessions) described in Tieman et al. (2017) (black). **(B)** Manhattan and **(B)** QQplot of the GWAS with differential growth during waterlogging as phenotype. Dashed gray and red lines in **(C)** indicate suggestive and significant -$\log_{10}$(P) values (-$\log_{10}$(P_{sug}) = 5.50 and -$\log_{10}$(P_{sign}) = 6.80). SNPs exceeding the suggestive threshold are colored red. Annotated genes are labeled with their corresponding Solyc ID.

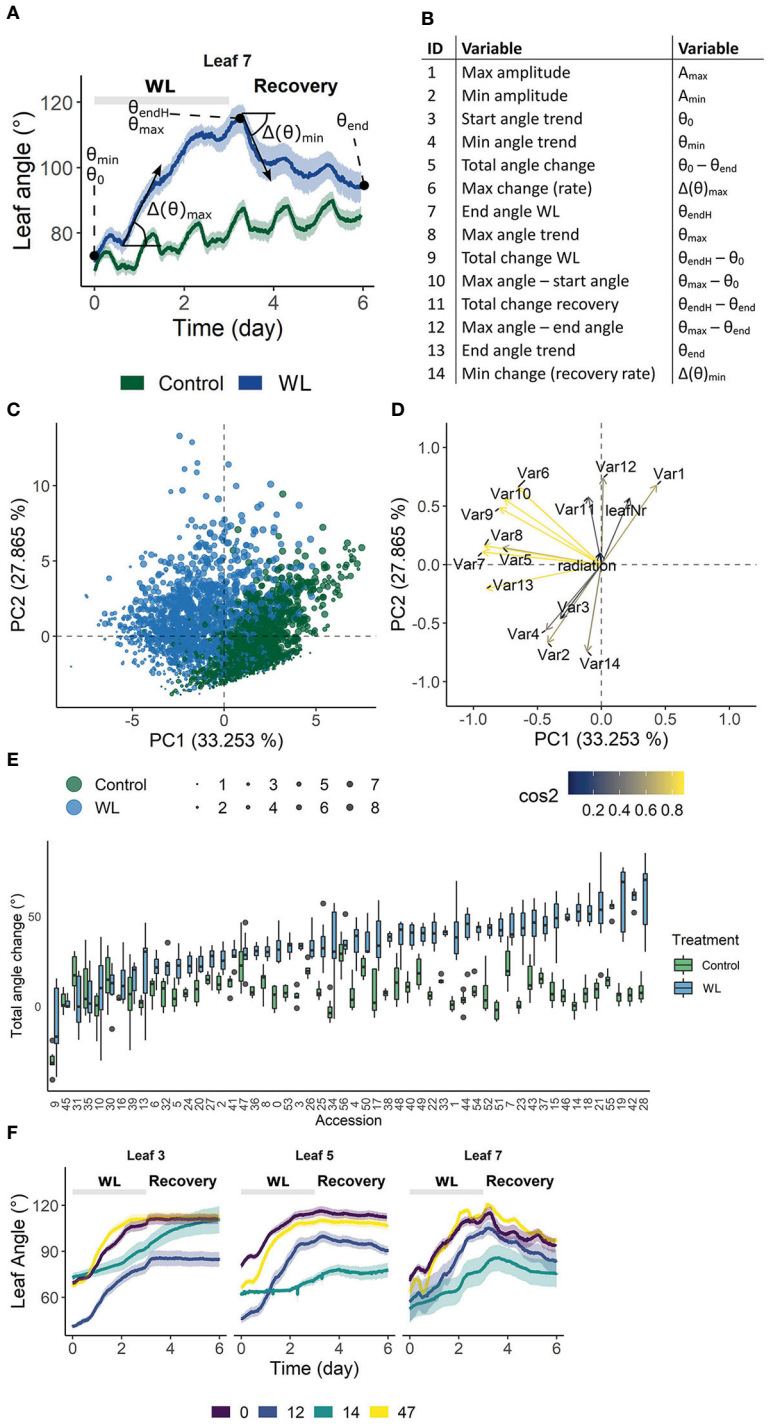
towards waterlogging stress in this species. To investigate the genetic basis of this mechanism, each of the angular traits was used separately as a phenotype for GWAS analysis, taking into account ontogeny. Besides these angular descriptors, we also fed the model with time-series angle data, revealing a number of critical points in the regulation of the epinastic response. At certain time points, both the count of associated SNPs and the rate of angular change induced by waterlogging were strongly elevated (Figures 4A, B). Both profiles also indicate that this regulation might be gated by the circadian clock, with major shifts caused by waterlogging starting mostly in the morning, following the oscillatory leaf movement. By analyzing the timing of divergence between waterlogging-treated plants and control plants, we quantified the onset of the epinastic response in Ailsa Craig, as a reference threshold for early responses. After 400 – 500 minutes (6 – 8 h), the rate of angular change increased significantly in leaves of waterlogged Ailsa Craig plants, providing a good reference time point for early epinastic responses in the GWAS.

Next, we filtered the SNP targets, retaining only SNPs with a –$\log_{10}$(P) value above the suggestive P-value and located in non-inflated chromosomic regions. These regions were determined by PCA on genome-wide P-test statistics, only retaining chromosomic regions with a limited total count of SNPs and –$\log_{10}$(P) (non-outliers in 5 subsequent PCAs). Only SNPs that could be annotated were further analyzed (Supplemental Table S4), revealing a grouping structure based on leaf age (lower panel Figure 4C) and angle dynamics (upper panel Figure 4C). SNPs associated with fast responses (< 10 h) were more enriched in mature leaves (leaf 3 – 5), and some of these SNPs were also detected during recovery (> 72 h). SNPs associated with angular responses in young leaves were mostly detected during the later stages of the treatment (> 34 h), corresponding to the second period of maximal divergence of angular rate between waterlogging and control (Figure 4A).

Specific gene groups are associated with early and late angular responses

In general, several SNPs and genes reoccurred throughout GWAS analyses for different phenotypes, and were associated with similar variables or subsequent time points in the time-series analysis. Given this clear time-dependent SNP profile (Figures 4B, C), we decided to tie SNP occurrence to distinct phases in the epinastic response (early: < 12 h; late: > 34 and < 72 h; recovery phase). While no genes harboring significant SNPs (higher than significant -$\log_{10}$(P)) were shared between these three phases, this was the case when including suggestive SNPs, indicating certain processes might still be shared (Supplemental Figure S4; Supplemental Tables S5-S9). To gain more insight in the early angle response, we focused on SNPs that were only detected at the

10.3389/fpls.2023.1178778



ID	Variable	Variable
1	Max amplitude	A_{max}
2	Min amplitude	A_{min}
3	Start angle trend	θ_0
4	Min angle trend	θ_{min}
5	Total angle change	$\theta_0 - \theta_{end}$
6	Max change (rate)	$\Delta(\theta)_{max}$
7	End angle WL	θ_{endH}
8	Max angle trend	θ_{max}
9	Total change WL	$\theta_{endH} - \theta_0$
10	Max angle – start angle	$\theta_{max} - \theta_0$
11	Total change recovery	$\theta_{endH} - \theta_{end}$
12	Max angle – end angle	$\theta_{max} - \theta_{end}$
13	End angle trend	θ_{end}
14	Min change (recovery rate)	$\Delta(\theta)_{min}$

FIGURE 3
Phenotypic variation of leaf angle traits during waterlogging and recovery of tomato. **(A, B)** Visualization of the 14 angle descriptors used in this study. **(C, D)** PCA on the angle descriptors showing the effect of waterlogging (treatment) and leaf development (leaf number) on leaf posture. **(C)** Observation plot of individual leaves and **(D)** variable plot of the descriptors defined in **(A, B)** with their quality of representation. **(E)** Natural variation in total leaf angle change ($\theta_{end} - \theta_0$) in control conditions and after 3 days of waterlogging and subsequent recovery of leaf 5. **(F)** Leaf angle dynamics of leaves of different ages (leaf 3, 5 and 7) of 4 different accessions. Colored bands represent the 95% confidence interval.

onset of the waterlogging treatment (< 12 h), possibly including key activators of epinasty. Most genes harboring one or more of these SNPs also reoccurred at later time-points and were mainly associated with responses for leaf 3 and 4. A small-scale gene ontology (GO) analysis showed that genes involved in plasma membrane processes (9 genes), defense responses (4 genes), signal transduction (4 genes) and two vacuolar (2 + 3 genes) and RNA processing (3 + 2 genes) categories were relatively abundant. Furthermore, a group of protein serine/threonine kinases and receptor like kinases was well-represented in this set (5 genes).

10.3389/fpls.2023.1178778

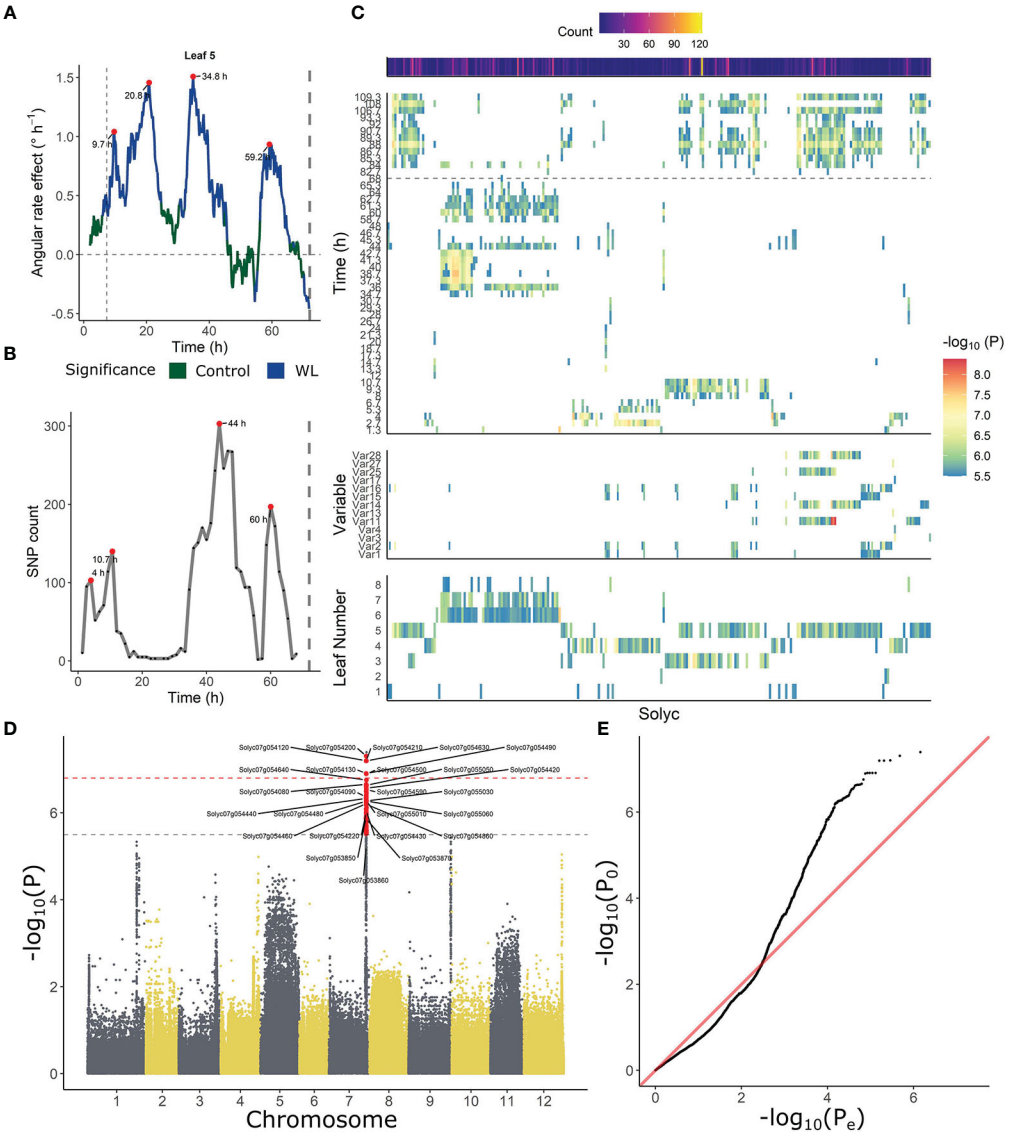


FIGURE 4
Association between leaf angle dynamics and SNP data during waterlogging and recovery in tomato. **(A, B)** Representation of **(A)** the effect of waterlogging on rate of angular change in leaf 5 and **(B)** SNP count, highlighting (red dots) certain periods of higher angular activity and higher number of associated SNPs. **(C)** Ontogenic (leaf 1 – 8) and time-dependent profile of annotated genes (Solyc number) associated with leaf angle changes during waterlogging. The heat maps depict the occurrence of significant associations of a certain gene in time (upper), with an angle variable (middle) and with leaf age (lower). **(D, E)** Manhattan and **(D)** QQplot of the GWAS analysis with leaf angle around a pivotal point (10.7 h) as phenotype (leaf 3). Dashed gray and red lines in **(D)** indicate suggestive and significant -$\log_{10}$(P) values (-$\log_{10}$(P_{sug}) = 5.50 and -$\log_{10}$(P_{sign}) = 6.80). SNPs exceeding the suggestive threshold are colored red. Annotated genes are labeled with their corresponding Solyc ID.

We further investigated genes with SNPs having a –$\log_{10}$(P-value) above the significance threshold and associated with early angle changes (< 12 h) (24 genes; Supplemental Table S5). The most reoccurring genes harboring SNPs of all analyzes combined were Cc-nbs-lrr, resistance protein (Solyc04g007050), Genomic DNA chromosome 3 P1 clone MYA6 (Solyc03g120490), S-receptor kinase-like protein 1 (Solyc07g005110), Vacuolar ATPase subunit H protein (Solyc07g005940), Ribonuclease 3-like protein 3 (Solyc07g005030) and Response regulator 23 (RR23) (Solyc07g005140). Other significant targets were involved in RNA processing, transport (e.g. ABC transporter C family member 2 (ABCC2)) and energy metabolism (e.g. Aldehyde dehydrogenase (ALDH)).

Besides these genes with SNPs specifically associated with early angular responses, other early-response genes reoccurred during the reoxygenation phase (when using suggestive SNPs), indicating they might also be involved in leaf repositioning. Genes discovered within the early waterlogging and recovery phase included a LRR receptor-like serine/threonine-protein kinase (EMS1-like, Solyc07g054120, 76 SNPs), Receptor-like kinase (SRF4, Solyc07g054500, 11 SNPs), a gene involved in energy metabolism (ATP synthase I-like, Solyc07g055050, 17 SNPs) and an unknown protein (Solyc07g054130, 22 SNPs). Several of these SNPs were associated with both early responses in older leaves (leaf 3) and late responses in younger leaves (leaf 5), indicating the corresponding

genes are ontogenically regulated in time. Genes present in both of these subsets included Transmembrane protein 34 (Solyc07g054320, 142 SNPs), B3 domain-containing protein (homolog of RELATED TO VERNALIZATION1 (AtRTV1), Solyc07g054630, 48 SNPs) and a NADH dehydrogenase (Solyc01g088350, 40 SNPs).

Leaf movement in young leaves is not limited to repositioning during reoxygenation. A set of 3 genes was specifically and significantly associated with movement in young leaves, mostly around 36 h after the start of the waterlogging treatment. This group consisted of a Polyamine oxidase (PO4, Solyc03g031880, 45 SNPs), a Coatomer protein epsilon subunit family protein (Solyc02g069590, 22 SNPs), and a LOB domain protein 4 (LBD4, Solyc02g069440, 10 SNPs).

Another 37 genes were present both during the late response (> 34 h) and during subsequent recovery, indicating they might be involved in late epinastic bending or stress recovery. There was no clear grouping of these individual genes, suggesting they might regulate distinct processes. However, multiple genes were involved in nuclear processing (6), including light-dependent short hypocotyls 1 (Solyc02g069510, 13 SNPs), a histone deacetylase (Solyc09g091440 22 SNPs) and a MYB85 transcription factor (Solyc07g054980, 10 SNPs). Also, a large number of SNPs exceeded the suggestive P-value, almost reaching the significant threshold, including Homeobox-leucine zipper protein AtHB14 (Solyc03g031760, 9 SNPs) and Auxin response factor 1-2 (ARF8B, Solyc03g031970, 7 SNPs).

A group of SNPs specifically associated with the recovery phase located to genes mainly enriched in membrane-bound receptor kinases. Besides these signaling components, the analysis also picked up Glucose transporter 8 (Solyc12g089180) during the reoxygenation phase. Other targets were mainly involved in ATP (8) and protein binding (6) and in nuclear processes (7), including histone modification (Solyc02g079300, Solyc09g091440 and Solyc01g008120) and transcription (e.g. Histidine kinase/DNA gyrase B (Solyc09g092100), MYB85 (Solyc07g054980)). This set also contained another aldehyde oxidase gene (AO4 pseudogene, Solyc01g088170, 66 SNPs) and two genes involved in specialized metabolism (Cinnamoyl CoA reductase-like protein (PAR1, Solyc01g008530) and Hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyl transferase (Solyc02g079490)).

Identification of potential regulatory targets

Next, we selected a few candidate target genes based on their timing and possible developmental association, to further study their involvement in leaf angle regulation (Table 1). Therefore, we verified if these genes were asymmetrically expressed within the petiole during waterlogging. Their gene expression level was quantified in both the abaxial and adaxial side of the petiole of leaf 5 after 48 h of waterlogging (Figure 5). Most of the selected targets showed only mild but non-significant changes. The lack of differences between control and epinastic petioles, is possibly due to sampling constraints and sample characteristics (e.g. containing a relative large proportion of vascular tissue). On the other hand, one of the several receptor-like kinases associated with the epinastic response at different time points (EMS-like1) did show a major shift during waterlogging. This LRR receptor-like serine/threonine-protein kinase, RLP (Solyc07g054120), harbored a large number of significant SNPs, especially at 560 minutes (9.33 h) after the start of the treatment in leaf 3 and also reoccurred during recovery in leaf 5. This gene was strongly downregulated in both petiole sections after 48 h of waterlogging (Figure 5).

Discussion

Waterlogging tolerance is genetically encoded in tomato

Genetic diversity has often been exploited as a natural resource for beneficial traits in breeding programs, including enhanced yield and stress resilience. For example, natural populations of Arabidopsis accessions have been screened for survival during drought (Kalladan et al., 2017) and flooding (Vashisht et al., 2011; Meng et al., 2022). However, insights in natural differentiation of waterlogging responses in crops such as tomato are mostly lacking. We have now shown that different tomato accessions display distinct growth responses when opposed to waterlogging and identified accessions with an enhanced recovery potential (Figure 1). We hypothesize that there are several survival strategies based on waterlogging resilience and tolerance. Some accessions wield a pause-strategy and display stagnant growth during waterlogging, followed by enhanced growth recovery during reoxygenation. Other accessions continue to grow during

TABLE 1 Target gene selection for RT-qPCR.

Solyc ID	Gene ID	Alternative annotation	Timing	Leaf
Solyc07g005140	RR23		Early	Mature (4)
Solyc07g054120	LRR receptor-like serine/threonine-protein kinase, RLP	EMS-like 1	Early Recovery	Old (3) Mature (5)
Solyc02g069440	LBD4		Late	Young (6 – 7)
Solyc02g079400	Nitric oxide synthase interacting protein (NOSIP)	CSU1	Recovery	Old – Young
Solyc07g054500	Receptor like kinase, RLK	SRF3/4	Early Recovery	Old (3) Mature (5)

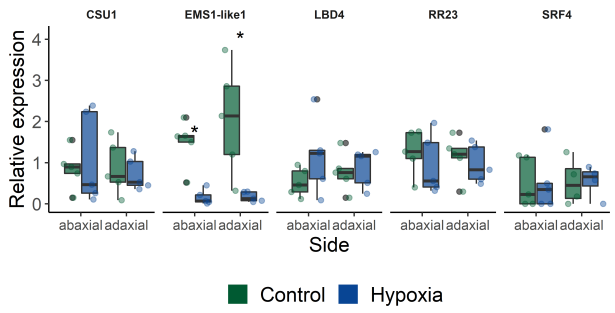


FIGURE 5
Expression of a subset of target genes identified in the GWAS (CSU1 = COP1 SUPPRESSOR 1; EMS1-like 1 = EXCESS MICROSPOROCYTES1-LIKE 1; LBD4 = LOB DOMAIN PROTEIN 4; RR23 = RESPONSE REGULATOR 23; SRF4 = STRUBBELIG-receptor family 4). Gene expression was quantified in abaxial and adaxial petiole sections (leaf 5) of plants subjected to waterlogging for 48 h (n = 5). Statistical significant differences ($P < 0.05$) are indicated with an asterisk.

waterlogging and maintain a similar growth rate afterwards, while other accessions undergo a strong growth reduction during both phases.

Our GWAS analysis revealed that a reduced growth rate during waterlogging and recovery were associated with a limited number of genes (Figure 2), including a MATH-BTB/POZ domain containing protein (SlBTB12, Solyc05g052960). It has been shown that AtBPM3, the closest homolog of SlBTB12 in Arabidopsis, can interact with the ethylene response factor APETALA2/ERF (AP2/ERF) transcription factor family members RAP2.4b and d (Weber and Hellmann, 2009; Rudnik et al., 2017), involved in light and ethylene signaling and overall plant development (Lin et al., 2008). Through its interaction with RAP2.4, BPM3 modulates growth and drought responses, independently of ABA (Lin et al., 2008). In contrast, through degradation of AtHB6, BPM3 can also regulate plant growth and interfere with ABA signaling (Lechner et al., 2011). In tomato, expression of SlBTB12 is controlled by various hormones and environmental conditions (Li et al., 2018), indicating this transcription factor might integrate signals for growth and abiotic stress such as waterlogging.

The second gene identified in this analysis, a hydroxycinnamoyl transferase (HCT) (Solyc05g052680), is part of the phenylpropanoid pathway. The HCT clade belongs to a larger family of BAHD acyltransferases, involved in the specialized metabolism, including the production of phenolic compounds such as anthocyanin and suberin in roots (Gou et al., 2009; Bontpart et al., 2015; Molina and Kosma, 2015). In tea (*Camellia sinensis*), overexpression of one of these BAHD genes affects not only the specialized metabolism, but also plant growth (Aktar et al., 2022). How and if this *HCT* gene aids tomato plants in waterlogging resilience remains to be investigated.

Leaf epinasty and metabolic changes are coregulated during waterlogging in tomato

The downwards bending of leaves (epinasty) reduces canopy cover, thereby limiting transpirational water losses (Else et al., 1995) and possible deleterious effects of photo-inhibition when carboxylation has ceased (Pastenes et al., 2005; Van Geest et al., 2012) during waterlogging. As such, epinasty and its recovery are part of the morphological adaptation to increase survival under low-oxygen stress conditions. We found a large genetic diversity in this waterlogging-induced epinastic bending, with some accessions showing a strong response while others displayed a mild or no response (Figure 3). The analysis also showed that shifts in angular change coincided with SNP frequency, defining key phases during waterlogging-induced epinastic bending and subsequent recovery.

Our GWAS analysis revealed a number of candidate genes that might be associated to this epinastic response, either by directly affecting leaf posture or by regulating other low-oxygen activated pathways (Figure 4). Several of these genes (e.g. Aldehyde dehydrogenase (Solyc07g005390); Amine oxidase family protein (Solyc03g031880)) were also retrieved in a transcriptomic study that assessed waterlogging of tomato (De Ollas et al., 2021), indicating they might be transcriptionally regulated during stress adaptation. Waterlogging-induced changes include rewiring of the carbohydrate and energy metabolism and the energy harvesting mechanism (Supplemental Table S10). For example, we identified SNPs in genes that include an aldehyde dehydrogenase (Solyc07g005390), beta-galactosidase (Solyc09g092160), Phosphoenolpyruvate carboxykinase (Solyc12g088160) and protochlorophyllide reductase like (Solyc07g054210). Recently, it has been shown that levels of the phytotoxic chlorophyll precursor protochlorophyllide are controlled by oxygen sensing in etiolated seedlings of tomato (Abbas et al., 2022).

Not only does this change indicate reprogramming of the energy metabolism, but also the detoxification of its ROS byproducts. Our GWAS detected SNPs in a thioredoxin reductase (Solyc03g032000), involved in ROS scavenging, and an aldehyde oxidase/xanthine dehydrogenase (AO/XDH) module potentially regulating ROS production and ABA biosynthesis (Sagi et al., 1999) (Solyc01g088200, Solyc01g088210, Solyc01g088230), indicating the need for ROS detoxification during waterlogging and recovery. Whether or not this directly affects leaf bending in tomato remains to be investigated, but XDH activity has been linked with leaf curling in Arabidopsis treated with the synthetic auxin 2,4-D (Pazmiño et al., 2014).

Besides apparent changes in the energy metabolism, our analyses also revealed some ambiguous associations related to genetic regulation, including DNA and RNA processing, small RNA synthesis and epigenetic regulation (Supplemental Table S10). This is not unexpected as plants need to activate a suite of adaptation responses to ensure their survival under waterlogging conditions. Especially epigenetic changes during stress, including chromatin remodeling, are gaining attention as they can determine the plasticity to respond adequately to changes in environmental conditions. Epigenetic changes have been observed in a range of species during flooding (Reynoso et al., 2019) and drought stress (Reynoso et al., 2022).

Is a network of kinases involved in low-oxygen stress signaling?

The variety of morphological and metabolic changes during waterlogging are the result of fine-tuned signal transduction

pathways, conveying messages within and between cells and over long distances between tissues. To our surprise, we found several SNPs in target genes involved in signaling and transport. Notably, we could identify several receptor-like kinases (RLK) and serine/threonine-protein kinases (12), including multiple G-type lectin S-receptor-like kinases (5) (Teixeira et al., 2018) and members of the wall-associated kinase (WAK) family (2) (Sun et al., 2020). Some of these targets have previously been reported to act in early and systemic signaling events in roots. For example, STRUBBELIG RECEPTOR KINASE 3 (SRF3), the closest Arabidopsis homolog to RLK (Solyc07g054500), coordinates root growth through iron sensing, while SERINE/THREONINE PROTEIN KINASE 3 (PK3/WAG1, homolog to Solyc06g069330) belongs to the PINOID family and modulates PIN polarity to guide auxin fluxes and direct root growth (Dhonukshe et al., 2015). We further identified 2 LRR containing proteins, including RLK (Solyc07g054120; EXCESS MICROSPOROCYTES1 (EMS1)-like, ITAG4.0), potentially involved in BR signaling. These SNP associations indicate that the activation of signaling cascades in roots is important to evoke waterlogging responses, possibly translating to epinastic bending.

Hormonal crosstalk during waterlogging-induced epinasty

In the past, it has been shown that waterlogging activates an array of short and long-distance hormonal signaling cascades. In our extended set of GWAS targets, we found several genes directly or indirectly involved in hormone biosynthesis and signaling pathways of ethylene, IAA, cytokinin (CK) and ABA (Supplemental Table S12). Ethylene, originating from root-borne ACC transport, is the master regulator of the epinastic response during waterlogging (Jackson and Campbell, 1976; English et al., 1995). During the early phase of epinastic bending, our analysis detected a SNP in the ethylene signaling gene Ethylene responsive transcription factor 2A (ERF2A; Solyc07g054220), exceeding the suggestive P-value for leaf 3. Downstream of ethylene, asymmetric growth in the petiole is likely mediated by local auxin redistribution (Lee et al., 2008). The homolog of AUXIN RESPONSE FACTOR 8A (ARF8A; Solyc03g031970), is controlled by the auxin/indole-3-acetic acid gene *SlIAA3/SHY2*, which is known to directly regulate epinastic bending in tomato treated with ethylene (Chaabouni et al., 2009). While the role of IAA3 in signal polarity seems to be developmentally encoded (Koyama et al., 2010), both AtIAA3/SHY2 and AtARF8 also respond to multiple hormonal cues (Brenner et al., 2005; Chaabouni et al., 2009). CK for example, directly downregulates *AtARF8* expression in Arabidopsis seedlings (Brenner et al., 2005). In addition, CK and BR signaling, together with changes in *SlIAA3* expression seem to be intertwined in regulating leaf posture in tomato (Li et al., 2016; Xia et al., 2021). While the role of ABA as regulator of stress responses, for example during waterlogging, has been described before (De Ollas et al., 2021), its impact on waterlogging-induced epinasty is unknown. Changes in ABA content have been previously linked with elongation in flooded *R. palustris* (Benschop et al., 2007) and rice (*Oryza sativa*) (Saika et al., 2007).

Besides the well-known classical phytohormones, other signaling compounds with hormone-like properties are gaining more attention. We identified a number of genes in our GWAS analysis which are involved in polyamine (PA) (Amine oxidase family protein Solyc03g031880; Ornithine carbamoyl transferase Solyc12g089210) and melatonin (Aromatic amino acid decarboxylase Solyc07g054280) biosynthesis. Both PA and melatonin have been shown to regulate plant growth and development and abiotic stress responses (reviewed in Chen et al. (2019) and Hardeland (2016)), making them potential mediators of waterlogging responses (Hurng et al., 1994). In tomato, PA metabolism and ethylene biosynthesis are closely intertwined, as both pathways use the same precursor S-adenosyl-L-methionine (Bellés et al., 1992; Van de Poel et al., 2013; Takács et al., 2021).

Light and hormonal signaling converge during waterlogging

Adequate and fast responses to stress conditions do not always require dedicated signaling pathways to enable survival. For example, plants repurpose the photomorphogenic machinery to activate escape strategies during flooding (van Veen et al., 2013). It is known that AtIAA3 (and AtARF8) combines light and auxin signaling to establish growth (Mao et al., 2020; Xi et al., 2021) in Arabidopsis, while SlIAA3 regulates ethylene-induced leaf bending in tomato (Chaabouni et al., 2009) and is upregulated during waterlogging (De Ollas et al., 2021). We found SNPs in at least three additional targets, known to operate on the interface between light and hormone signaling (Light-dependent short hypocotyls 1 (LSH1) Solyc02g069510; a GH3 family protein Solyc07g054580, homolog to DWARF IN LIGHT 2 (AtDFL2); Nitric oxide synthase interacting protein, homolog to AtCSU1). In Arabidopsis LSH1 (Zhao et al., 2004) and DFL2 (Takase et al., 2003) integrate red and blue light signals to regulate seedling growth, while CSU1 targets CONSTITUTIVE PHOTOMORPHOGENIC1 (COP1) in dark-grown seedlings (Xu et al., 2014). The possible role of these targets in regulating growth or responses such as epinasty during waterlogging is unknown.

Petiole polarity as a regulator of asymmetric responses

The involvement of signaling pathways in regulating leaf epinasty eventually require some level of polarity to direct asymmetric growth of the petiole to facilitate downwards bending. At least two target genes (LBD4, Solyc02g069440; Homeobox-leucine zipper protein AtHB14, Solyc03g031760) derived from our GWAS could be related to this process. AtLBD4 is a member of a family of LOB-domain containing proteins, specifically expressed at organ boundaries (Shuai et al., 2002). AtLBD4 itself seems to act at the phloem-procambium interface to regulate vascular development (Smit et al., 2020) and secondary growth in roots downstream of CK signaling (Ye et al., 2021). In *Medicago truncatula*, another LBD transcription factor, ELONGATED PETIOLULE 1 (ELP1), has been shown to define motor organ identity and thus regulate leaf movement (Chen et al., 2012).

Also the formation of the dorsoventral axis during early leaf development is tightly controlled by leaf polarity. The homeobox-leucine zipper protein (Solyc03g031760) identified in our study has not yet been properly annotated nor classified. However, it has resemblance with HOMEODOMAIN GLABROUS 2 (HDG2)-like or MERISTEM LAYER 1 (AtML1) in Arabidopsis, involved in epidermal differentiation. In tomato, it seems to coincide with the LANATA locus, responsible for trichome development (Xie et al., 2022).

Conclusion

Through the integration of longitudinal leaf angle data in a GWAS, we were able to devise a dynamic profile of SNPs related to waterlogging-induced epinasty in tomato. This approach allowed us to identify target genes that are at least associated to the angular response. Whether or not these genes are causal or rather part of the general waterlogging response remains to be investigated. Nevertheless, the timing of the above SNP profile showed resemblance to changes in the angular rate during waterlogging, providing a framework for both initiation and maintenance of the epinastic response. Despite the broad range of gene classes, certain of these targets could be functionally involved in morphological responses and resilience to waterlogging. We identified genes involved in energy and carbohydrate metabolism and hormone signaling which needs to be rewired to sustain plant survival during stress. These dynamic transitions allow the cessation and reactivation of growth during and after a waterlogging event. By regulating the above processes, different tomato accessions seem to invest in different strategies to ensure waterlogging survival.

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

BG and BV designed the experiments, JP and BG phenotyped the tomato accessions, BG analyzed the data, BG performed qPCR analysis and BG and BV wrote the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was funded by the Research Foundation Flanders with a FWO PhD fellowship (11C4319N; 1150822N) to BG and JP respectively, and a FWO research grant (G092419N) to BVdP; by a KU Leuven research grant (C14/18/056) to BVdP and a PhD-back-up grant (DB/17/007/BM) to BG. This work was also established in the framework of the RoxyCost action of the EU (CA18210).

Acknowledgments

We thank the KU Leuven Greenhouse Core Facility for assistance in plant cultivation.

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1178778/full#supplementary-material

References

Abbas, M., Sharma, G., Dambire, C., Marquez, J., Alonso-Blanco, C., Proaño, K., et al. (2022). An oxygen-sensing mechanism for angiosperm adaptation to altitude. *Nature* 606 (7914), 565–569. doi: 10.1038/s41586-022-04740-y

Aktar, S., Bai, P., Wang, L., Xun, H., Zhang, R., Wu, L., et al. (2022). Identification of a BAHD acyltransferase gene involved in plant growth and secondary metabolism in tea plants. *Plants (Basel)* 11 (19), 2483. doi: 10.3390/plants11192483

Bellés, J. M., Tornero, P., and Conejero, V. (1992). Pathogenesis-related proteins and polyamines in a developmental mutant of tomato, epinastic. *Plant Physiol.* 98 (4), 1502–1505. doi: 10.1104/pp.98.4.1502

Benschop, J. J., Millenaar, F. F., Smeets, M. E., Van Zanten, M., Voesenek, L. A. C. J., and Peeters, A. J. M. (2007). Abscisic acid antagonizes ethylene-induced hyponastic growth in arabidopsis. *Plant Physiol.* 143 (2), 1013–1023. doi: 10.1104/pp.106.092700

Bontpart, T., Cheynier, V., Ageorges, A., and Terrier, N. (2015). BAHD or SCPL acyltransferase? what a dilemma for acylation in the world of plant phenolic compounds. *New Phytol.* 208 (3), 695–707. doi: 10.1111/nph.13498

Brenner, W. G., Romanov, G. A., Köllmer, I., Bürkle, L., and Schmülling, T. (2005). Immediate-early and delayed cytokinin response genes of arabidopsis thaliana identified by genome-wide expression profiling reveal novel cytokinin-sensitive processes and suggest cytokinin action through transcriptional cascades. *Plant J.* 44 (2), 314–333. doi: 10.1111/j.1365-313X.2005.02530.x

Chaabouni, S., Jones, B., Delalande, C., Wang, H., Li, Z., Mila, I., et al. (2009). Sl-IAA3, a tomato Aux/IAA at the crossroads of auxin and ethylene signalling involved in differential growth. *J. Exp. Bot.* 60 (4), 1349–1362. doi: 10.1093/jxb/erp009

Chen, J., Moreau, C., Liu, Y., Kawaguchi, M., Hofer, J., Ellis, N., et al. (2012). Conserved genetic determinant of motor organ identity in medicago truncatula and related legumes. *Proc. Natl. Acad. Sci. United States America* 109 (29), 11723–11728. doi: 10.1073/pnas.1204566109

Chen, D., Shao, Q., Yin, L., Younis, A., and Zheng, B. (2019). Polyamine function in plants: metabolism, regulation on development, and roles in abiotic stress responses. *Front. Plant Sci.* 9 (January). doi: 10.3389/fpls.2018.01945

Cox, M. C. H., Benschop, J. J., Vreeburg, R. A. M., Wagemaker, C. A. M., Moritz, T., Peeters, A. J. M., et al. (2004). The roles of ethylene, auxin, abscisic acid, and gibberellin in the hyponastic growth of submerged rumex palustris petioles. *Plant Physiol.* 136 (2), 2948–2960. doi: 10.1104/pp.104.049197

De Ollas, C., González-Guzmán, M., Pitarch, Z., Matus, J. T., Candela, H., Rambla, J. L., et al. (2021). Identification of ABA-mediated genetic and metabolic responses to soil flooding in tomato (Solanum lycopersicum l. mill). *Front. Plant Sci.* 12 (March). doi: 10.3389/fpls.2021.613059

Dhonukshe, P., Huang, F., Galvan-Ampudia, C. S., Mähönen, A. P., Kleinevehn, J., Xu, J., et al. (2015). Plasma membrane-bound AGC3 kinases phosphorylate PIN auxin carriers at TPRXS(N/S) motifs to direct apical PIN recycling. *Dev. (Cambridge)* 142 (13), 2386–2387. doi: 10.1242/dev.127415

Else, M. A., Davies, W. J., Malone, M., and Jackson, M. B. (1995). A negative hydraulic message from oxygen-deficient roots of tomato plants? (Influence of soil flooding on leaf water potential, leaf expansion, and synchrony between stomatal conductance and root hydraulic conductivity). *Plant Physiol.* 109 (3), 1017–1024. doi: 10.1104/pp.109.3.1017

English, P. J., Lycett, G. W., Roberts, J. A., and Jackson, M. B. (1995). Increased 1-Aminocyclopropane-1-Carboxylic acid oxidase activity in shoots of flooded tomato plants raises ethylene production to physiologically active levels. *Plant Physiol.* 109 (4), 1435–1440. doi: 10.1104/pp.109.4.1435

Geldhof, B., Pattyn, J., Eyland, D., Carpentier, S., and Van de Poel, B. (2021). A digital sensor to measure real-time leaf movements and detect abiotic stress in plants. *Plant Physiol.* 187 (3), 1131–1148. doi: 10.1093/plphys/kiab407

Gou, J. Y., Yu, X. H., and Liu, C. J. (2009). A hydroxycinnamoyltransferase responsible for synthesizing suberin aromatics in arabidopsis. *Proc. Natl. Acad. Sci. United States America* 106 (44), 18855–18860. doi: 10.1073/pnas.0905555106

Hardeland, R. (2016). Melatonin in plants – diversity of levels and multiplicity of functions. *Front. Plant Sci.* 7 (FEB2016). doi: 10.3389/fpls.2016.00198

Hebelstrup, K. H., van Zanten, M., Mandon, J., Voesenek, L. A. C. J., Harren, F. J. M., Cristescu, S. M., et al. (2012). Haemoglobin modulates NO emission and hyponasty under hypoxia-related stress in arabidopsis thaliana. *J. Exp. Bot.* 63 (15), 5581–5591. doi: 10.1093/jxb/ers210

Hurng, W. P., Lur, H. S., Liao, C. K., and Kao, C. H. (1994). Role of abscisic acid, ethylene and polyamines in flooding-promoted senescence of tobacco leaves. *J. Plant Physiol.* 143 (1), 102–105. doi: 10.1016/S0176-1617(11)82104-8

Jackson, M. B., and Campbell, D. J. (1976). Waterlogging and petiole epinasty in tomato: the role of ethylene and low oxygen. *New Phytol.* 76 (1), 21–29. doi: 10.1111/j.1469-8137.1976.tb01434.x

Kalladan, R., Lasky, J. R., Chang, T. Z., Sharma, S., Juenger, T. E., and Verslues, P. E. (2017). Natural variation identifies genes affecting droughtinduced abscisic acid accumulation in arabidopsis thaliana. *Proc. Natl. Acad. Sci. United States America* 114 (43), 11536–11541. doi: 10.1073/pnas.1705884114

Kang, H. M., Sul, J. H., Service, S. K., Zaitlen, N. A., Kong, S. Y., Freimer, N. B., et al. (2010). Variance component model to account for sample structure in genome-wide association studies. *Nat. Genet.* 42 (4), 348–354. doi: 10.1038/ng.548

Koyama, T., Mitsuda, N., Seki, M., Shinozaki, K., and Ohme-Takagi, M. (2010). TCP Transcription factors regulate the activities of ASYMMETRIC LEAVES1 and miR164, as well as the auxin response, during differentiation of leaves in arabidopsis. *Plant Cell* 22 (11), 3574–3588. doi: 10.1105/tpc.110.075598

Lechner, E., Leonhardt, N., Eisler, H., Parmentier, Y., Alioua, M., Jacquet, H., et al. (2011). MATH/BTB CRL3 receptors target the homeodomain-leucine zipper ATHB6 to modulate abscisic acid signaling. *Dev. Cell* 21 (6), 1116–1128. doi: 10.1016/j.devcel.2011.10.018

Lee, Y., Jung, J. W., Kim, S. K., Hwang, Y. S., Lee, J. S., and Kim, S. H. (2008). Ethylene-induced opposite redistributions of calcium and auxin are essential components in the development of tomato petiolar epinastic curvature. *Plant Physiol. Biochem.* 46 (7), 685–693. doi: 10.1016/j.plaphy.2008.04.003

Li, X. J., Chen, X. J., Guo, X., Yin, L. L., Ahammed, G. J., Xu, C. J., et al. (2016). DWARF overexpression induces alteration in phytohormone homeostasis, development, architecture and carotenoid accumulation in tomato. *Plant Biotechnol. J.* 14 (3), 1021–1033. doi: 10.1111/pbi.12474

Li, J., Su, X., Wang, Y., Yang, W., Pan, Y., Su, C., et al. (2018). Genome-wide identification and expression analysis of the BTB domain-containing protein gene family in tomato. *Genes Genomics* 40 (1), 1–15. doi: 10.1007/s13258-017-0604-x

Lin, R. C., Park, H. J., and Wang, H. Y. (2008). Role of arabidopsis RAP2.4 in regulating lightand ethylene-mediated developmental processes and drought stress tolerance. *Mol. Plant* 1 (1), 42–57. doi: 10.1093/mp/ssm004

Mantilla-Perez, M. B., Bao, Y., Tang, L., Schnable, P. S., and Salas-Fernandez, M. G. (2020). Toward "Smart canopy" sorghum: discovery of the genetic control of leaf angle across layers. *Plant Physiol.* 184 (4), 1927–1940. doi: 10.1104/pp.20.00632

Mantilla-Perez, M. B., and Salas Fernandez, M. G. (2017). Differential manipulation of leaf angle throughout the canopy: current status and prospects. *J. Exp. Bot.* 68 (21–22), 5699–5717. doi: 10.1093/jxb/erx378

Mao, Z., He, S., Xu, F., Wei, X., Jiang, L., Liu, Y., et al. (2020). Photoexcited CRY1 and phyB interact directly with ARF6 and ARF8 to regulate their DNA-binding activity and auxin-induced hypocotyl elongation in arabidopsis. *New Phytol.* 225 (2), 848–865. doi: 10.1111/nph.16194

Meng, X., Li, L., Pascual, J., Rahikainen, M., Yi, C., Jost, R., et al. (2022). GWAS on multiple traits identifies mitochondrial ACONITASE3 as important for acclimation to submergence stress. *Plant Physiol.* 188 (4), 2039–2058. doi: 10.1093/plphys/kiac011

Molina, I., and Kosma, D. (2015). Role of HXXXD-motif/BAHD acyltransferases in the biosynthesis of extracellular lipids. *Plant Cell Rep.* 34 (4), 587–601. doi: 10.1007/s00299-014-1721-5

Olson, D. C., Oetiker, J. H., and Yang, S. F. (1995). Analysis of LE-ACS3, a 1-Aminocyclopropane-1-carboxylic acid synthase gene expressed during flooding in the roots of tomato plants. *J. Biol. Chem.* 270 (23), 14056–14061. doi: 10.1074/jbc.270.23.14056

Pantazopoulou, C. K., Bongers, F. J., Küpers, J. J., Reinen, E., Das, D., Evers, J. B., et al. (2017). Neighbor detection at the leaf tip adaptively regulates upward leaf movement through spatial auxin dynamics. *Proc. Natl. Acad. Sci. United States America* 114 (28), 7450–7455. doi: 10.1073/pnas.1702275114

Park, Y. J., Lee, H. J., Gil, K. E., Kim, J. Y., Lee, J. H., Lee, H., et al. (2019). Developmental programming of thermonastic leaf movement. *Plant Physiol.* 180 (2), 1185–1197. doi: 10.1104/pp.19.00139

Pastenes, C., Pimentel, P., and Lillo, J. (2005). Leaf movements and photoinhibition in relation to water stress in field-grown beans. *J. Exp. Bot.* 56 (411), 425–433. doi: 10.1093/jxb/eri061

Pazmiño, D. M., Rodríguez-Serrano, M., Sanz, M., Romero-Puertas, M. C., and Sandalio, L. M. (2014). Regulation of epinasty induced by 2,4-dichlorophenoxyacetic acid in pea and arabidopsis plants. *Plant Biol.* 16 (4), 809–818. doi: 10.1111/plb.12128

Polko, J. K., Pierik, R., van Zanten, M., Tarkowská, D., Strnad, M., Voesenek, L. A. C. J., 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

Polko, J. K., van Zanten, M., van Rooij, J. A., Marée, A. F. M., Voesenek, L. A. C. J., Peeters, A. J. M., et al. (2012). Ethylene-induced differential petiole growth in arabidopsis thaliana involves local microtubule reorientation and cell expansion. *New Phytol.* 193 (2), 339–348. doi: 10.1111/j.1469-8137.2011.03920.x

Rauf, M., Arif, M., Fisahn, J., Xue, G. P., Balazadeh, S., and Mueller-Roeber, B. (2013). NAC transcription factor SPEEDY HYPONASTIC GROWTH regulates flooding-induced leaf movement in arabidopsis. *Plant Cell* 25 (12), 4941–4955. doi: 10.1105/tpc.113.117861

R Core Team (2019). *R: a language and environment for statistical computing* (Vienna, Austria: R Foundation for Statistical Computing). Available at: https://www.R-project.org/

Reynoso, M. A., Borowsky, A. T., Pauluzzi, G. C., Yeung, E., Zhang, J., Formentin, E., et al. (2022). Gene regulatory networks shape developmental plasticity of root cell types under water extremes in rice. *Dev. Cell* 57 (9), 1177–1192.e6. doi: 10.1016/j.devcel.2022.04.013

Reynoso, M. A., Kajala, K., Bajic, M., West, D. A., Pauluzzi, G., Yao, A. I., et al. (2019). Evolutionary flexibility in flooding response circuitry in angiosperms. *Science* 365 (6459), 1291–1295. doi: 10.1126/science.aax8862

Rudnik, R., Bulcha, J. T., Reifschneider, E., Ellersiek, U., and Baier, M. (2017). Specificity versus redundancy in the RAP2.4 transcription factor family of arabidopsis thaliana: transcriptional regulation of genes for chloroplast peroxidases. *BMC Plant Biol.* 17 (1), 1–17. doi: 10.1186/s12870-017-1092-5

Sagi, M., Fluhr, R., and Lips, S. H. (1999). Aldehyde oxidase and xanthine dehydrogenase in a flacca tomato mutant with deficient abscisic acid and wilty phenotype. *Plant Physiol.* 120 (2), 571–577. doi: 10.1104/pp.120.2.571

Saika, H., Okamoto, M., Miyoshi, K., Kushiro, T., Shinoda, S., Jikumaru, Y., et al. (2007). Ethylene promotes submergence-induced expression of OsABA8ox1, a gene that encodes ABA 8′-hydroxylase in rice. *Plant Cell Physiol.* 48 (2), 287–298. doi: 10.1093/pcp/pcm003

Shiu, O. Y., Oetiker, J. H., Yip, W. K., and Yang, S. F. A. (1998). The promoter of LE-ACS7, an early flooding-induced 1-aminocyclopropane-1-carboxylate synthase gene of the tomato, is tagged by a Sol3 transposon. *Proc. Natl. Acad. Sci. United States America* 95 (17), 10334–10339. doi: 10.1073/pnas.95.17.10334

Shuai, B., Reynaga-Peña, C. G., and Springer, P. S. (2002). The lateral organ boundaries gene defines a novel, plant-specific gene family. *Plant Physiol.* 129 (2), 747–761. doi: 10.1104/pp.010926

Smit, M. E., McGregor, S. R., Sun, H., Gough, C., Bågman, A. M., Soyars, C. L., et al. (2020). A PXY-mediated transcriptional network integrates signaling mechanisms to control vascular development in arabidopsis. *Plant Cell* 32 (2), 319–335. doi: 10.1105/tpc.19.00562

Sun, Z., Song, Y., Chen, D., Zang, Y., Zhang, Q., Yi, Y., et al. (2020). Genome-wide identification, classification, characterization, and expression analysis of the wall-associated kinase family during fruit development and under wound stress in tomato (Solanum lycopersicum l.). *Genes* 11 (10), 1–20. doi: 10.3390/genes11101186

Takács, Z., Poór, P., and Tari, I. (2021). Interaction between polyamines and ethylene in the response to salicylic acid under normal photoperiod and prolonged darkness. *Plant Physiol. Biochem.* 167 (August), 470–480. doi: 10.1016/j.plaphy.2021.08.009

Takase, T., Nakazawa, M., Ishikawa, A., Manabe, K., and Matsui, M. (2003). DFL2, a new member of the arabidopsis GH3 gene family, is involved in red light-specific hypocotyl elongation. *Plant Cell Physiol.* 44 (10), 1071–1080. doi: 10.1093/pcp/pcg130

Teixeira, M. A., Rajewski, A., He, J., Castaneda, O. G., Litt, A., and Kaloshian, I. (2018). Classification and phylogenetic analyses of the arabidopsis and tomato G-type lectin receptor kinases. *BMC Genomics* 19 (1), 1–20. doi: 10.1186/s12864-018-4606-0

Tieman, D., Zhu, G., Resende, M. F. R., Lin, T., Nguyen, C., Bies, D., et al. (2017). A chemical genetic roadmap to improved tomato flavor. *Science* 355 (6323), 391–394. doi: 10.1126/science.aal1556

van de Poel, B., Bulens, I., Markoula, A., Hertog, M. L. A. T. M., Dreesen, R., Wirtz, M., et al. (2012). Targeted systems biology profiling of tomato fruit reveals coordination of the yang cycle and a distinct regulation of ethylene biosynthesis during postclimacteric ripening. *Plant Physiol.* 160 (3), 1498–1514. doi: 10.1104/pp.112.206086

Van de Poel, B., Bulens, I., Oppermann, Y., Hertog, M. L. A. T. M., Nicolai, B. M., Sauter, M., et al. (2013). S-adenosyl-l-methionine usage during climacteric ripening of tomato in relation to ethylene and polyamine biosynthesis and transmethylation capacity. *Physiologia Plantarum* 148 (2), 176–188. doi: 10.1111/j.1399-3054.2012.01703.x

Van Geest, G., Van Ieperen, W., Post, A. G., and Schoutsen, C. G. L. M. (2012). Leaf epinasty in chrysanthemum: enabling breeding against an adverse trait by physiological research. *Acta Hortic.* 953, 345–350.

van Veen, H., Mustroph, A., Barding, G. A., Vergeer-van Eijk, M., Welschen-Evertman, R. A. M., Pedersen, O., et al. (2013). Two rumex species from contrasting hydrological niches regulate flooding tolerance through distinct mechanisms. *Plant Cell* 25 (11), 4691–4707. doi: 10.1105/tpc.113.119016

Vashisht, D., Hesselink, A., Pierik, R., Ammerlaan, J. M. H., Bailey-Serres, J., Visser, E. J. W., et al. (2011). Natural variation of submergence tolerance among arabidopsis thaliana accessions. *New Phytol.* 190 (2), 299–310. doi: 10.1111/j.1469-8137.2010.03552.x

Weber, H., and Hellmann, H. (2009). Arabidopsis thaliana BTB/POZ-MATH proteins interact with members of the ERF/AP2 transcription factor family. *FEBS J.* 276 (22), 6624–6635. doi: 10.1111/j.1742-4658.2009.07373.x

Wickham, H. (2009). ggplot2: Elegant Graphics for Data Analysis (Use R!). (New York: Springer).

Xi, Y., Yang, Y., Yang, J., Zhang, X., Pan, Y., and Guo, H. (2021). IAA3-mediated repression of PIF proteins coordinates light and auxin signaling in arabidopsis. *PloS Genet.* 17 (2), 1–20. doi: 10.1371/JOURNAL.PGEN.1009384

Xia, X., Dong, H., Yin, Y., Song, X., Gu, X., Sang, K., et al. (2021). Brassinosteroid signaling integrates multiple pathways to release apical dominance in tomato. *PNAS* 118 (11), e2004384118. doi: 10.1073/pnas.2004384118

Xie, Q., Xiong, C., Yang, Q., Zheng, F., Larkin, R. M., Zhang, J., et al. (2022). A novel regulatory complex mediated by lanata (ln) controls multicellular trichome formation in tomato. *New Phytol* 236, 2294–2310. doi: 10.1111/nph.18492

Xu, D., Lin, F., Jiang, Y., Huang, X., Li, J., Ling, J., et al. (2014). The RING-finger E3 ubiquitin ligase COP1 SUPPRESSOR1 negatively regulates COP1 abundance in maintaining COP1 homeostasis in dark-grown arabidopsis seedlings. *Plant Cell* 26 (5), 1981–1991. doi: 10.1105/tpc.114.124024

Ye, L., Wang, X., Lyu, M., Siligato, R., Eswaran, G., Vainio, L., et al. (2021). Cytokinins initiate secondary growth in the arabidopsis root through a set of LBD genes. *Curr. Biol.* 31 (15), 3365–3373.e7. doi: 10.1016/j.cub.2021.05.036

Zhao, L., Nakazawa, M., Takase, T., Manabe, K., Kobayashi, M., Seki, M., et al. (2004). Overexpression of LSH1, a member of an uncharacterised gene family, causes enhanced light regulation of seedling development. *Plant J.* 37 (5), 694–706. doi: 10.1111/j.1365-313X.2003.01993.x

Zhou, Y., Bhattacharjee, S., Carroll, C., Chen, Y., Dai, X., Fan, J., et al. (2022). *Fdapace: functional data analysis and empirical dynamics*. Available at: https://cran.r-project.org/web/packages/fdapace/index.html (accessed March 2022).

Zhou, X., and Stephens, M. (2014). Efficient multivariate linear mixed model algorithms for genome-wide association studies. *Nat. Methods* 11 (4), 407–409. doi: 10.1038/nmeth.2848

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 01 August 2023
DOI 10.3389/fpls.2023.1153040

OPEN ACCESS

EDITED BY
Baohong Zhang,
East Carolina University, United States

REVIEWED BY
Jiban Shrestha,
Nepal Agricultural Research Council, Nepal
Dinesh Kumar Saini,
South Dakota State University, United States
Zitong Li,
Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australia
Yiwen Wang,
Chinese Academy of Agricultural Sciences, China

*CORRESPONDENCE
Carlos Maldonado
cmaldo1782@gmail.com

RECEIVED 28 January 2023
ACCEPTED 12 July 2023
PUBLISHED 01 August 2023

CITATION
Mora-Poblete F, Maldonado C, Henrique L, Uhdre R, Scapim CA and Mangolim CA (2023) Multi-trait and multi-environment genomic prediction for flowering traits in maize: a deep learning approach. *Front. Plant Sci.* 14:1153040. doi: 10.3389/fpls.2023.1153040

Multi-trait and multi-environment genomic prediction for flowering traits in maize: a deep learning approach

Freddy Mora-Poblete[1], Carlos Maldonado[2]*, Luma Henrique[3], Renan Uhdre[3], Carlos Alberto Scapim[3] and Claudete Aparecida Mangolim[4]

[1]Institute of Biological Sciences, University of Talca, Talca, Chile, [2]Centro de Genómica y Bioinformática, Facultad de Ciencias, Universidad Mayor, Santiago, Chile, [3]Department of Agronomy, State University of Maringá, Paraná, Brazil, [4]Department of Biotechnology, Genetics and Cell Biology, State University of Maringá, Paraná, Brazil

Maize (*Zea mays* L.), the third most widely cultivated cereal crop in the world, plays a critical role in global food security. To improve the efficiency of selecting superior genotypes in breeding programs, researchers have aimed to identify key genomic regions that impact agronomic traits. In this study, the performance of multi-trait, multi-environment deep learning models was compared to that of Bayesian models (Markov Chain Monte Carlo generalized linear mixed models (MCMCglmm), Bayesian Genomic Genotype-Environment Interaction (BGGE), and Bayesian Multi-Trait and Multi-Environment (BMTME)) in terms of the prediction accuracy of flowering-related traits (Anthesis-Silking Interval: ASI, Female Flowering: FF, and Male Flowering: MF). A tropical maize panel of 258 inbred lines from Brazil was evaluated in three sites (Cambira-2018, Sabaudia-2018, and Iguatemi-2020 and 2021) using approximately 290,000 single nucleotide polymorphisms (SNPs). The results demonstrated a 14.4% increase in prediction accuracy when employing multi-trait models compared to the use of a single trait in a single environment approach. The accuracy of predictions also improved by 6.4% when using a single trait in a multi-environment scheme compared to using multi-trait analysis. Additionally, deep learning models consistently outperformed Bayesian models in both single and multiple trait and environment approaches. A complementary genome-wide association study identified associations with 26 candidate genes related to flowering time traits, and 31 marker-trait associations were identified, accounting for 37%, 37%, and 22% of the phenotypic variation of ASI, FF and MF, respectively. In conclusion, our findings suggest that deep learning models have the potential to significantly improve the accuracy of predictions, regardless of the approach used and provide support for the efficacy of this method in genomic selection for flowering-related traits in tropical maize.

KEYWORDS

Bayesian models, deep learning, multi-trait, multi-environment, genomic prediction, candidate genes

1 Introduction

Maize (*Zea mays* L.) is a crucial cereal crop that plays a vital role in global food security, biofuel production, and animal feed (Maldonado et al., 2020; Grote et al., 2021). Consumed by over 4.5 billion people, particularly in rural areas of Latin America and Africa, it is an important source of calories and nutrients (Domínguez-Hernández et al., 2022). With its high genetic diversity and ease of sexual reproduction, maize is a versatile crop that offers many agronomic and reproductive advantages. Its separate inflorescences (male and female) also allow for easily controlled crosses and the creation of highly inbred lines – high levels of genetic homozygosity in the lines – (Strable and Scanlon, 2009). As a result of these advantages, variousss agronomic traits such as grain yield, flowering time, and nutritional value have been improved through breeding programs worldwide (e.g., Alves et al., 2018; Gedil and Menkir, 2019).

Flowering time is an agriculturally important trait for crop production that can be manipulated by various approaches such as breeding and genetic modifications (Hirohata et al., 2022). In maize, it has been shown that flowering time is significantly associated with regional adaptation and is a complex trait controlled by hundreds of loci with small effects, many with multiple allelic series (Romero et al., 2017). The genetic control of flowering time involves networks of genes that interact with environmental conditions, which is a determining factor in the duration of the crop cycle (Parent et al., 2018). Conventional approaches in quantitative genetics, such as QTL (Quantitative Trait *loci*) mapping, genomic selection, and genome-wide association studies (GWAS), have traditionally been used to investigate the genetic basis of the quantitative variation in flowering time-related traits. For example, Romero et al. (2017) assayed the potential for predicting flowering time in maize landraces using GBLUP (Genomic Best Linear Unbiased Predictor); a widely used statistical method for genomic selection. Across trials, the average fivefold cross-validated prediction accuracy was 0.45 for flowering time using either 30,000 markers or one SNP for each of the most significant genes. Similarly, Maldonado et al. (2020) used deep learning predictive models and found a predictive ability of up to 0.78 for maize traits related to flowering.

Other genetic studies in maize have emphasized the importance of identifying genetic variants (QTLs) controlling flowering time-related traits under a wide range of environmental conditions to improve stress tolerance (Leng et al., 2022). The study conducted by Maldonado et al. (2019) who used a population of inbred lines of tropical maize, identified a total of 45 SNPs and 44 Haplotype-block significantly associated with flowering time, which was distributed across the entire genome. Moreover, the study also found that some of the loci identified were associated with multiple flowering-related traits, which suggests a possible pleiotropic effect of these loci. Additionally, the study found that some loci displayed associations with multiple flowering-related traits. This observation suggests the presence of a potential pleiotropic effect, where a single genetic locus influences the expression of multiple traits related to flowering. Another study carried out by Birnbaum and Roberts (2019), which aimed to identify SNPs significantly associated with flowering time in a panel of maize inbred lines by using a GWAS approach, identified a total of 25 significant SNPs for flowering time, of which 15 were novel, and 10 were previously reported. The study also identified several candidate genes underlying the significant SNPs that were associated with flowering traits. Overall, these studies demonstrate that GWAS can provide valuable information for understanding the genetic basis of flowering time-related traits, which can inform the development of improved maize varieties.

Various studies have highlighted the potential of using GWAS and genomic selection approaches in enhancing crop breeding and developing improved maize varieties (e.g., Liu et al., 2021; Ma and Cao, 2021; Vinayan et al., 2021). For example, the study conducted by Zhou et al. (2021) aimed to identify the genetic variants associated with yield and yield-related traits in maize crops. The results of the study found that the combination of these methods provided the best results in predicting the breeding value of individuals for yield and yield-related traits. Furthermore, the study identified several loci associated with yield and yield-related traits, which demonstrate the effectiveness of the combined GWAS and genomic selection approach in identifying genetic variants associated with these traits. On the other hand, recent studies have placed significant emphasis on the advancement of more precise predictive models, such as multi-trait or multi-environment genomic prediction models. These models have shown remarkable improvements in prediction accuracy when compared to uni-trait models, especially when traits are correlated. Additionally, they have proven beneficial in predicting traits that are difficult or expensive to phenotype (Gill et al., 2021). As breeders routinely gather phenotypic data across numerous traits and diverse environments, extending the application of multi-trait approaches to incorporate genotype-by-environment interactions could further enhance the accuracy of genomic prediction models within breeding programs (Montesinos-López et al., 2019; Hu et al., 2022). Multi-trait and multi-environment Bayesian and Deep Learning models have been proposed by Montesinos-López et al. (2016) (Bayesian multi-trait and multi-environment; BMTME), Montesinos-López et al. (2018) (Deep learning multi-trait and multi-environment; DL), Granato et al., 2018 (Bayesian Genomic Genotype × Environment Interaction; BGGE) and Hadfield and Nakagawa (2010) (MCMC Generalised Linear Mixed Models; MCMCglmm). Sandhu et al. (2022) showed that the multi-trait DL approach improved the accuracy of genomic prediction compared to uni-trait and multi-trait+multi-environment (BMTME) models. This highlights the potential of using multi-trait, multi-environment deep learning models in genomic prediction and crop breeding. The study highlights the potential of using multi-trait, multi-environment deep learning models in genomic prediction and crop breeding.

Uni- and Multi-trait (UT and MT, respectively), as well as, Uni- and Multi-environment (UE and ME, respectively) approaches have been compared keeping fixed the traits (UTUE vs UTME, or MTUE vs MTME) or environments (UTUE vs MTUE, or UTME vs MTME) as one (Uni) or multiple (Multi). However, comparisons

among all approaches simultaneously have not been performed yet, particularly for traits exhibiting low or negative correlations. Thus, the present study aimed to evaluate the performance of these four approaches for the genomic prediction of flowering-related traits in tropical maize using the Bayesian and deep learning approaches. To accomplish this, a panel of 258 tropical maize inbred lines was analyzed using SNP markers. In addition, a complementary genome-wide association study, coupled with network-assisted gene prioritization (post-GWAS), was performed to identify potential candidate genes associated with these traits. The results of this study provide insights into the potential of using deep learning models for enhancing prediction accuracy in the context of genomic selection for flowering-related traits in tropical maize.

2 Materials and methods

2.1 Plant materials

The study utilized a panel of 258 tropical maize inbred lines from the core collection germplasm of the State University of Maringa, Parana State, Brazil, which were derived from three genetic backgrounds: field corn, popcorn, and sweet corn genotypes (Supplementary Table S1). Genomic prediction models were developed using phenotypic records derived from three locations within the state of Paraná, Brazil: Cambira, Sabaudia and Iguatemi, during the growing seasons of 2017-2018 (Cambira and Sabaudia), 2019-2020 (Iguatemi), and 2020-2021 (Iguatemi). Complementary, a genome-wide association study was performed using Iguatemi data (both growing seasons), and then, these results were compared with the other locations following the study by Maldonado et al. (2019).

2.2 Experimental design and trait measurement

The experimental design for Cambira and Sabaudia was an alpha-lattice with 24 incomplete blocks and 3 replications per line, while in Iguatemi, the lines were planted according to a partially balanced incomplete block design in a 17x17 square lattice with 4 replications per line. The following flowering-related traits were evaluated: Female Flowering time (FF) measured as the number of days from sowing to visible silks, Male Flowering time (MF) measured as the number of days from sowing to anther extrusion from the tassel glumes, and Anthesis-Silking Interval (ASI) calculated as the difference between MF and FF (Maldonado et al., 2019; Maldonado et al., 2020).

2.3 Phenotypic data analysis

The analysis of the phenotypic data was performed using the following Bayesian model available in the package "MCMCglmm" (Hadfield and Nakagawa, 2010) of R software (Team R. C, 2013):

$$y = X\beta + Zf + \epsilon \quad (1)$$

where y is the vector of the phenotypic observations, X and Z are the known incidence matrices that relate the observation vector (y) to the vectors β and f, respectively. β is the vector of replications and block within replications, f is the vector of family effects and ϵ is the vector of residuals or error vector. The *y* vector corresponds to the adjusted phenotypic observations, which were utilized in the subsequent sections for Genomic Prediction Models and Genome-Wide Association Study.

Correlations between each pair of traits were calculated using a Bayesian bi-trait model (MCMCglmm), according to Maldonado et al. (2019), using the following expression:

$$r_{xy} = \frac{\hat{\sigma}_{G_{xy}}}{\sqrt{\hat{\sigma}^2_{G_x} * \hat{\sigma}^2_{G_y}}} \quad (2)$$

where $\hat{\sigma}_{G_{xy}}$ correspond to posterior distribution samples of genotypic covariance between the traits, and $\hat{\sigma}^2_{G_x}$, $\hat{\sigma}^2_{G_y}$ correspond to posterior mean distribution samples of genotypic variance for each pair of traits under analysis.

2.4 Genotyping, population structure and linkage disequilibrium

Genomic DNA was extracted from the leaf tissue of 21-day-old plants using the protocol described by Maldonado et al. (2019), which follows the method developed by Chen and Ronald (1999). The DNA samples were then sent to the University of Wisconsin-Madison Biotechnology Center for SNP discovery through genotyping by sequencing (Elshire et al., 2011; Glaubitz et al., 2014). Monomorphic SNP markers and those with a call rate lower than 90% were removed, and SNPs with a minor allele frequency (MAF) of less than 0.05 were eliminated, resulting in 291,633 high-quality SNPs. Finally, missing data were imputed through linkage disequilibrium k-nearest neighbor imputation (Money et al., 2015), as described in Maldonado et al. (2020).

The kinship matrix was calculated using the identity-by-state method (Endelman and Jannink, 2012) with the TASSEL 5.2 software (Bradbury et al., 2007). The population genetic structure was inferred using a Bayesian clustering model in the InStruct 2.3.4 program (Gao et al., 2007). Ten runs were performed for each possible value of K (number of genetically differentiated groups), ranging from 1 to 6, with 100,000 Monte Carlo Markov Chain replicates and a burn-in period of 10,000 iterations. The optimal value of K was determined using the second-order change rate of the probability function with respect to K (ΔK), as proposed by Evanno et al. (2005) and the lowest deviance information criterion (DIC). Additionally, a t-distributed stochastic neighbor embedding (t-SNE) visualization was performed using Python 3.7 language and the Keras 2.2.4 and TensorFlow 1.14.0 libraries to corroborate the results from InStruct. A perplexity of 30, a learning rate of 200 and 1,000 iterations were used in the t-SNE model according to López-Cortés et al. (2020).

The Linkage Disequilibrium (LD) was estimated using the correlation coefficients of allelic frequencies (r2) calculated for all possible allele combinations. The critical r2 value was determined using the transformation of the square root of the r2 values as proposed by Breseghello and Sorrells (2006), with the 95th percentile of these data serving as the threshold.

2.5 Genomic prediction models and cross-validation

2.5.1 Markov chain Monte Carlo generalized linear mixed models

In this study, Uni-Trait-Uni-Environment and Multi-Trait-Uni-Environment analyses were implemented according to Mathew et al. (2016) and Torres et al. (2018). The Uni- and Multi-Trait approaches were implemented using the following model:

$$y_i = X_i\beta_i + Z_i u_i + \epsilon_i, \qquad i = 1,\ 2\ldots,\ n \tag{3}$$

where y_i is the vector of the phenotypic values of the traits, β_i and u_i are vectors of fixed and random effects associated with trait i, respectively, and ϵi is a vector of error terms, which are independently normally distributed with mean zero and variance σ_ϵ^2. Moreover, X_i and Z_i are incidence matrices for the fixed and random effects for trait i, respectively. Then mixed model equation (MME) for the above model is:

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{bmatrix} \begin{bmatrix} \beta \\ u \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix} \tag{4}$$

where R and G are covariance matrices associated with the vectors ϵ and u of residuals and random effects, respectively. If $R0$ is the residual covariance for more than one trait, then R can be calculated as $R=R0\otimes I$ ($\otimes$ represent the Kronecker product between $R0$ and the identity matrix). Similarly, the genetic covariance matrix G can be calculated as $G=G0\otimes A$, where A and $G0$ are the additive genetic relationship matrix and additive genetic (co)variance matrix, respectively. The MCMCglmm R package (Hadfield and Nakagawa, 2010; Team R. C, 2013) was used to implement the model, using 100,000 iterations, a 10,000 burn-in period, and a sampling interval of 5.

2.5.2 Bayesian genomic genotype × environment interaction

The Uni-Trait-Multi-Environment approach was implemented using the BGGE R package (Granato et al., 2018) within R software (Team R. C, 2013). This package utilizes Bayesian hierarchical modeling to solve linear mixed models, as described in Granato et al. (2018) and Costa-Neto et al. (2020), in which the distribution of the transformed data d, given b and σ_ϵ^2, is:

$$f(d|b,\sigma_\epsilon^2) = \prod_{i=1}^{n} N(d_i|b_i,\sigma_\epsilon^2) \tag{5}$$

The Bayesian linear mixed model assumes that $p(u|\sigma_u^2) = N(u|0, K\sigma_u^2)$; the conditional distribution of b_i is given as $p(b_i|\sigma_u^2) = N(b_i|0, K\sigma_u^2 s_i)$, where si is the eigenvalues. The BGGE package assumes that conjugate prior distribution of σ_u^2 and σ_ϵ^2 are given by inverse chi-squared with $p(\sigma_\epsilon^2) \sim \chi^{-2}(\nu_u, Sc_u)$ and $p(\sigma_\epsilon^2) \sim \chi^{-2}(\nu_\epsilon, Sc_\epsilon)$, respectively, in which which ν_u and ν_ϵ denote the degree of freedom, and Sc_u and Sc_ϵ the scale factors for μ and ϵ. Then, the joint posterior distribution of $(b, \sigma_u^2, \sigma_\epsilon^2)$, given d, ν_μ, Sc_u, ν_ϵ, Sc_ϵ and S, is:

$$\begin{aligned} &p(b, \sigma_u^2, \sigma_\epsilon^2|d, \nu_u, Sc_u, \nu_\epsilon, Sc_\epsilon, S) \\ &\propto \left\{ \prod_{i=1}^{n} N(d_i|b_i, \sigma_\epsilon^2)\ \ N(b_i|0, \sigma_u^2 s_i) \right\} \times \chi^{-2}(\sigma_u^2|\nu_u, \nu_u Sc_u) \\ &\times \chi^{-2}(\sigma_\epsilon^2|\nu_\epsilon, \nu_\epsilon Sc_\epsilon) \end{aligned} \tag{6}$$

Finally, the BGGE analysis was conducted using 100,000 iterations, with a 10,000 iteration burn-in period and a thinning of 5.

2.5.3 Bayesian multi-trait and multi-environment

The Multi-Trait-Multi-Environment analysis was carried out using the BMTME R package (Montesinos-López et al., 2016) within R software (Team R. C, 2013). The BMTME model is defined as (Montesinos-López et al., 2018; Sandhu et al., 2022):

$$y = X\beta + Z_1 b_1 + Z_2 b_2 + \epsilon \tag{7}$$

where y is the matrix of order t x l, with t is the number of traits and $l = e$ x g, where e and g are the numbers of environments and genotypes, respectively; X, Z_1, and Z_2 are design matrixes for environmental effect, genotypic effect, and genotype by environmental interaction, respectively; β is beta coefficient matrix of order e x t; b_1 is the random genotypic effect of genotype × trait interaction distributed as $b_1 \sim$ MN(0, G, Σt), where G is additive relationship matrix calculated using the VanRaden (2008) and Σt is the unstructured covariance matrix of order t x t; b_2 is the random genotypic x trait x environment effect matrix distributed as $b_2 \sim$ MN(0, Σe G, Σt), where Σe is the unstructured covariance matrix of order e x e. BMTME was performed considering 10,000 burn-in and 100,000 test iterations.

2.5.4 Uni- and multi-trait, uni- and multi-environment deep learning

In this study, Deep Learning methods were used to analyze Uni- and Multi-Trait, Uni- and Multi-Environment data, as described in Montesinos-López et al. (2018); Crossa et al. (2019) and Montesinos-López et al. (2019). A densely connected network was chosen as it does not assume a specific structure for the input features. This network typically includes an input layer, T output layers (for multi-trait modeling), and hidden layers between the input and output layers. This type of neural network is commonly referred to as a feedforward neural network (Figure 1).

In this study, we employed a neural network architecture with multiple layers to predict flowering traits in tropical maize (Figure 1). The network consists of an input layer with "n" neurons, representing the number of features in the dataset. Following the input layer, three hidden layers were incorporated, each containing 50 neurons. These hidden neurons perform non-

10.3389/fpls.2023.1153040

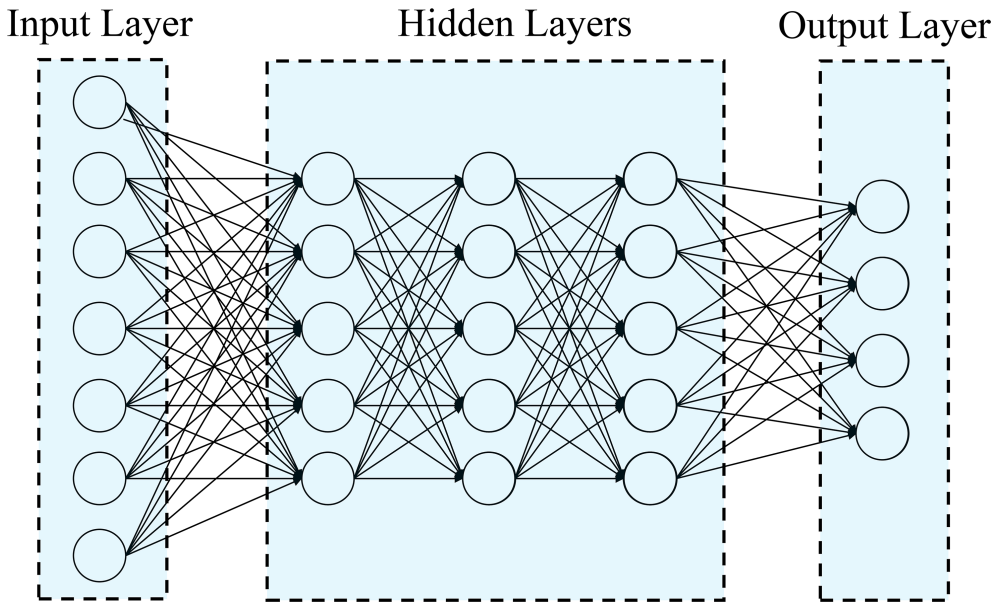


FIGURE 1
Example of feedforward deep neural network with one input layer (with n neurons that correspond to the input information), three hidden layers (each layer with M neurons) and one output layer (with o neurons that correspond to number of traits to be predicted).

linear transformations on the original input attributes, as described by Montesinos-López et al. (2018). For the output layer, the network was designed to have one neuron for uni-trait predictions and four neurons for multi-trait predictions. The number of output neurons corresponds to the number of response variables we aimed to predict for flowering traits. The neurons in the network are fully connected, and the strength of the connection weights determines the contribution of each neuron to the overall network output. A regularization technique known as dropout was implemented to temporarily removes a random subset of neurons and their connections during the training process, enhancing the network's ability to generalize and avoid overfitting (Montesinos-López et al., 2019).

The analytical forms of the model depicted in Figure 1 can be represented by the following equation (Montesinos-López et al., 2019):

$$V = g\left(\sum_{p=1}^{N} w_{jp} V_p\right) \quad \text{for hidden layers} \tag{8}$$

$$y_o = g\left(\sum_{p=1}^{N} w_{op} V_p\right) \quad \text{for output layer} \tag{9}$$

where N denotes the total number of input variables in each layer, w_{jp} and w_{jp} represents the weight of the input in hidden (with j=1, …, M neurons) and output (with o=1, …, O neurons) layers, respectively, while V_p represents the value of the pth input variable, and g represents the activation function. In this network, each layer generates the output for each neuron in the subsequent layer, ultimately producing the output for each response variable of interest. The learning process involves adjusting the weights that connect the layers to optimize the model's performance. The input variables for the multi-trait approach corresponded to the concatenation of environments, markers through the Cholesky decomposition of the genomic relationship matrix, and genotype × environment interaction ($G\times E$). For this purpose, the design matrices of environments (ZE), genotypes (ZG) and $G\times E$ (ZGE) were built, followed by the Cholesky decomposition of the genomic relationship matrix (G). Then, the design matrix of genotypes was post-multiplied by the transpose of the upper triangular factor of the Cholesky decomposition (QT), $Z_G^{\star} = Z_G Q^T$, followed by the calculation of the $G\times E$ term as the product of the design matrix of the $G\times E$ term post-multiplied by the Kronecker product of the identity matrix of order equal to the number of environments and QT, that is, $Z_{GE}^{\star} = Z_{GE}(I_I \otimes Q^T)$. After that, the matrix with input covariates used for implementing Deep Learning models was equal to $X = [Z_E, Z_G^{\star}, Z_{GE}^{\star}]$. It should be noted that Uni-Trait approach uses the same implementation as the multi-trait approach described above but with a feedforward neural network with only one neuron in the output layer.

In this study, deep learning models were implemented using the R code of Montesinos-López et al. (2018) in R software (Team R. C, 2013). The following hyperparameters were considered: 50 units (U), 200 epochs, 3 hidden layers, rectified linear activation unit (ReLU) as the activation function, and the dropout regularization method for training the models.

2.5.5 Cross validation

The genomic prediction methods were evaluated using four approaches: Uni-Trait-Uni-Environment (UTUE), Uni-Trait-Multi-Environment (MTUE), Multi-Trait-Uni-Environment (MTUE) and Multi-Trait-Multi-Environment (MTME). These approaches were tested in two scenarios: I) randomly selecting independent training (80%) and validation (20%) groups (for each

trait in each site), in which 50 cycles of cross-validation were performed, and II) predicting the second season of environment Iguatemi (validation dataset) using the first season of environment Iguatemi (training dataset DT1), other environments (Cambira and Sabaudia; training dataset DT2), and other environments (Cambira and Sabaudia) plus the first season of Iguatemi (training dataset DT3).

The prediction accuracy was evaluated by calculating the average Pearson correlation coefficient between the observed and predicted phenotypes in the validation set for all models (Deep Learning, MCMCglmm, BMTME, and BGGE).

2.6 GWAS, candidate genes and co-functional networks

The Genome-Wide Association Study (GWAS) was conducted using the mixed linear model (MLM) in TASSEL 5.2 (Bradbury et al., 2007) for the three flowering traits (FF, MF, and ASI). The statistical model incorporated the effects of population structure (Q) and genetic relationships or kinship matrix (K) among the inbred lines, as represented by the following mixed model:

$$y = S\alpha + Qv + Z\mu + \epsilon \quad (10)$$

where y is the vector of adjusted phenotypic observations, α and v are the vectors of fixed effects of molecular markers and population structure, respectively, μ and ϵ are the vectors of random effects of polygenic effects and residual, respectively. S, Q and Z are the incidence matrices of the associated vectors.

The probability of a locus being associated with two or more traits was evaluated using the Bayes Factor (BF) and Posterior Probability of Association (PPA) (Stephens and Balding, 2009). The PPA was calculated by considering the BF and prior probability of association, as outlined by Stephens and Balding (2009):

$$PPA = \frac{(BF \ x \ \pi)}{(1-\pi) + (BF \ x \ \pi)} \quad (11)$$

where π is the significance level of SNP associated with the trait of interest. BF was calculated using Bayesian multivariate regression analysis in the SNPTEST software (Marchini and Band, 2016) according to Maldonado et al. (2019).

The candidate genes surrounding the significant SNPs identified by GWAS were selected by establishing a window of twice the distance indicated by the LD around the SNP, with the SNP serving as the center of the window. These candidate genes were then prioritized using MaizeNet (Lee et al., 2019) by analyzing their connections to genes previously associated with flowering time in *Zea mays*. Co-functional networks were also constructed by linking the candidate genes to subnetworks enriched for gene ontology annotations related to biological processes involved in flowering.

GWAS, identified candidate genes, and constructed co-functional networks were applied for the Iguatemi, seasons 1 and 2. Results for the Cambira and Sabaudia environments can be found in the study by Maldonado et al. (2019).

3 Results

In this study, the genetic correlations between female flowering (FF) and male flowering (MF) remained consistent across all environments (Figure 2) with a positive correlation ($r > 0.82$) and highly significant ($p<0.001$). However, the correlation between the anthesis-silking interval (ASI) and the other two traits was inconsistent across environments, showing both positive and negative correlation values. Furthermore, the correlation of the flowering traits among the different environments (Cambira, Sabaudia, Iguatemi season 1, and Iguatemi season 2) were positive and statistically significant (Figure 3). Notably, MF had the highest correlations among the environments Cambira, Sabaudia and Iguatemi season 1, while FF had the highest correlation values among Iguatemi season 2 and other environments (Figure 3).

3.1 Genetic structure and linkage disequilibrium

In this study, a Bayesian clustering analysis was conducted on 258 tropical inbred lines, resulting in the grouping of these lines into two genetic clusters (as determined by the lowest DIC value and the highest ΔK). Cluster I and II consisted of 83 (with 82 popcorn and one field corn genotypes) and 175 maize lines (comprising 151 field corn, 13 popcorn, and all sweet corn lines) respectively.

The t-SNE method was used to visualize the SNP data, and it clearly separated the two clusters through its second dimension (t-SNE2), which was consistent with the results obtained from InStruct (Figure 4). The t-SNE method effectively maintained the distributions of the original data space (by matching pairwise similarity distributions) in a lower-dimensional projected space (Chan et al., 2018).

Linkage disequilibrium (LD) was also estimated at the genome-wide level and for each individual chromosome (Supplementary Table S2). The LD decayed rapidly within 2.7 kb, with a cut-off value of r2 = 0.12. Chromosomes 3 and 7 showed a faster LD decay than the other chromosomes, with values of about 2.12 kb and a cut-off of r2 = 0.12. Conversely, chromosome 4 presented the slowest LD decay, with a value of 5.35 kb.

3.2 Genome-wide association study

The results of the genome-wide association study (GWAS) for the flowering traits in Iguatemi seasons are presented in Table 1. A total of 31 SNPs were identified as being associated with the three traits of interest across both Iguatemi seasons, with 13 SNPs associated in the first season (Iguatemi 2020), 18 in the second season (Iguatemi 2021), and one in both seasons. Of these, 11 SNPs were associated with ASI, with 5 identified in the first season and 6 in the second season. Similarly, 11 SNPs were found to be associated with FF, with 5 identified in season 1 and 6 in season 2. Lastly, 9 SNPs were associated with FM, with 3 identified in season 1 and 6 in season 2. Notably, two SNPs (S5_217372319 and S6_150165479)

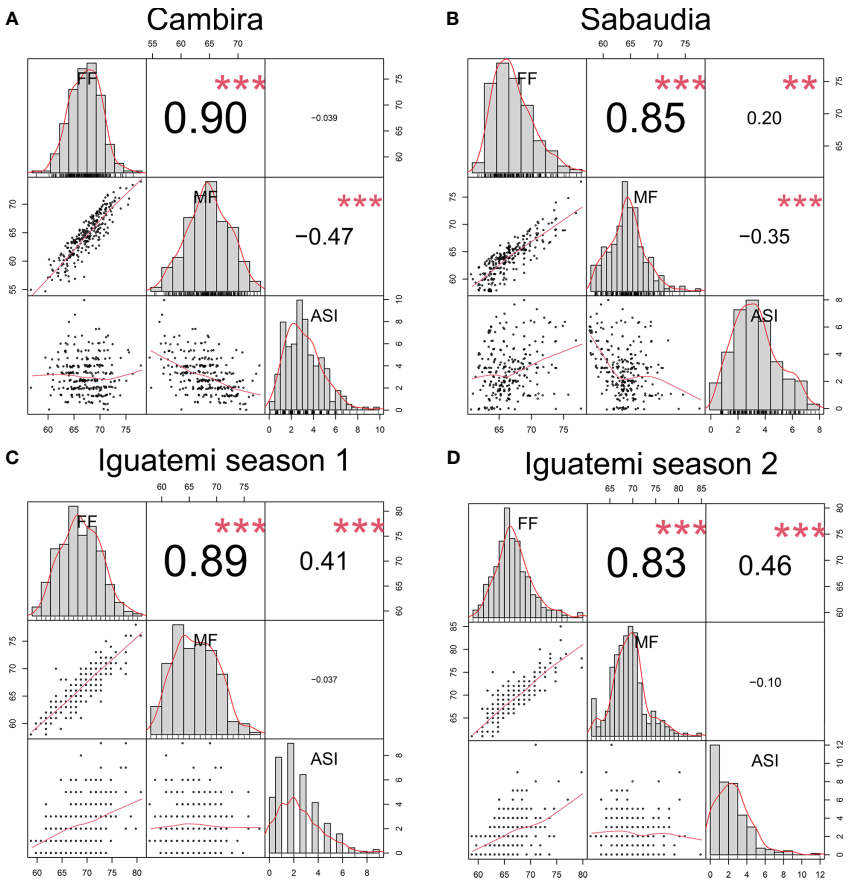


FIGURE 2
Correlation between flowering traits in the Cambira, Sabaudia, Iguatemi season 1, and Iguatemi season 2 environments (**A–D**, respectively). The figure illustrates the correlation between female flowering (FF), male flowering (FM), and anthesis-silking interval (ASI) in the four different environments. The diagonal of the plot displays histograms and distributions of the observed phenotype values, while the lower off-diagonal presents scatter plots between the traits. Significance levels of the correlation coefficients are indicated by ** for p< 0.01, and *** for p< 0.001.

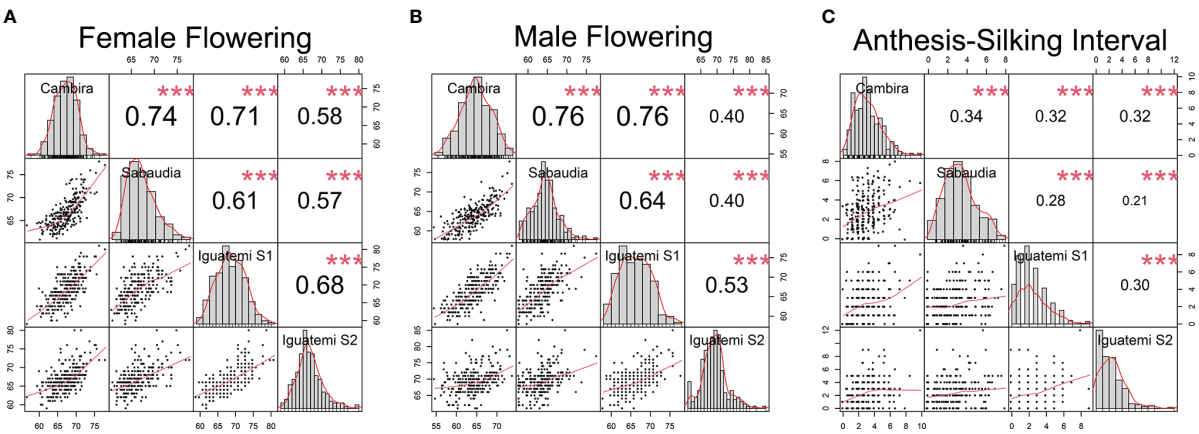


FIGURE 3
Correlation among the study environments (Cambira, Sabaudia, Iguatemi season 1 and Iguatemi season 2) for each flowering trait: female and male flowering (**A, B**, respectively); and anthesis-silking interval **(C)**. The diagonal line of the plot illustrates the histograms and the distribution of the observed phenotype values for each trait across all environments. The lower off-diagonal section presents the scatterplot between the environments for each trait, whereas the upper off-diagonal section displays the correlation coefficient between environments for each trait. Significance levels of the correlation coefficients is indicated by *** for p< 0.001.

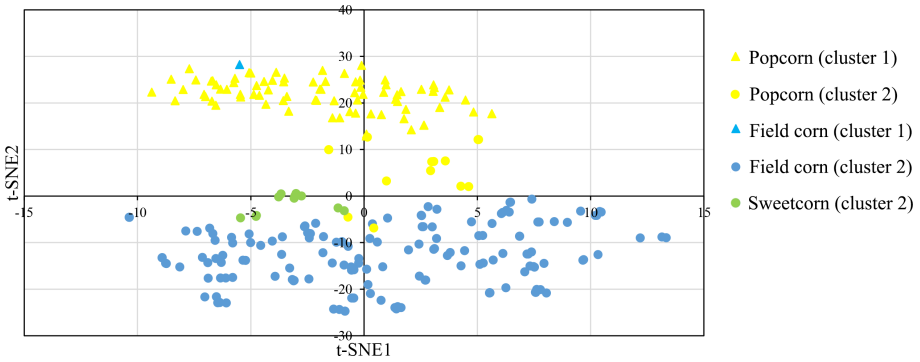


FIGURE 4
t-distributed stochastic neighbor embedding (t-SNE) visualization of the genetic relatedness of 258 maize inbred lines using a genome-wide panel of 291,633 SNP markers. The visualization is color-coded by population, with yellow representing Popcorn, blue representing Field corn, and green indicating Sweetcorn. The shapes of the individual points indicate an individual's proportion of ancestry to genetically differentiated groups determined by InStruct, with triangles indicating cluster 1 and circles indicating cluster 2.

were concomitantly associated with both FF and FM traits, suggesting a possible pleiotropic effect. To confirm this, multivariate Bayesian regression was performed on these loci in relation to the FF and FM traits. This analysis yielded PPA values of 0.99 and 0.74 for S6_150165479 and S5_217372319, respectively, and log10 (BF) > 5.1 for both loci, further supporting the pleiotropic effect of these loci as indicated in the Supplementary Table S3.

In season 1 (Iguatemi 2020), the proportion of the phenotypic variance (PV%) explained by SNP markers was 37%, 37%, and 22% of the phenotypic variation of ASI, FF, and FM, respectively (Table 1). In season 2 (Iguatemi 2021), the PV% explained for ASI and FM was higher than in the first season, at 49% and 45%, respectively.

3.3 Candidate genes and co-functional networks

A total of 26 candidate genes were identified based on the physical position of these SNPs in relation to the maize reference genome B73 (Supplementary Table S4). These candidate genes were found to be neighboring to the associated SNPs, with 12, seven and six candidate genes related to ASI, FF, and FM traits, respectively. Notably, four SNPs were located close to the same candidate genes, resulting in 22 unique candidate genes being identified in the present analysis.

The application of network-assisted prioritization using the MaizeNet database revealed 93 additional candidate genes

TABLE 1 Summary of the associations detected by a genome-wide association study for the traits of female/male flowering time (FF and MF, respectively) and anthesis–silking interval (ASI).

Trait	Iguatemi Season	Marker	Chr	Pos	p-value	PV%	BIN
ASI	1	S1_214720998	1	214720998	$5.53X10^{-06}$	7.9%	1.07
	1	S1_95747751	1	95747751	$5.56X10^{-06}$	7.8%	1.05
	1	S1_12340947	1	12340947	$1.40X10^{-05}$	7.1%	1.01
	1	S2_142739572	2	142739572	$1.46X10^{-05}$	7.2%	2.05
	1	S8_112412901	8	112412901	$1.86X10^{-05}$	7.1%	8.04
	2	S4_245982321	4	245982321	$7.40X10^{-07}$	9.6%	4.11
	2	S1_214720998	1	214720998	$5.43X10^{-06}$	8.0%	1.07
	2	S3_122398302	3	122398302	$7.91X10^{-06}$	7.8%	3.04
	2	S3_122398313	3	122398313	$7.91X10^{-06}$	7.8%	3.04
	2	S3_122398320	3	122398320	$7.91X10^{-06}$	7.8%	3.04
	2	S4_245029688	4	245029688	$9.45X10^{-06}$	7.6%	4.11
FF	1	S8_18474015	8	18474015	$2.65X10^{-06}$	8.0%	8.02

(Continued)

TABLE 1 Continued

Trait	Iguatemi Season	Marker	Chr	Pos	p-value	PV%	BIN
	1	S6_150165479	6	150165479	$5.00X10^{-06}$	7.5%	6.05
	1	S10_14797601	10	14797601	$8.82X10^{-06}$	7.2%	10.03
	1	S8_143046924	8	143046924	$9.44X10^{-06}$	7.2%	8.05
	1	S7_468747	7	468747	$1.74X10^{-05}$	6.7%	7.00
	2	S5_42052202	5	42052202	$3.89X10^{-06}$	7.8%	5.03
	2	S5_217372319	5	217372319	$1.82X10^{-05}$	7.6%	5.09
	2	S2_15002111	2	15002111	$1.11X10^{-05}$	7.2%	2.02
	2	S2_47411894	2	47411894	$1.27X10^{-05}$	7.0%	2.04
	2	S2_43966599	2	43966599	$1.40X10^{-05}$	6.9%	2.04
	2	S7_34488540	7	34488540	$1.40X10^{-05}$	7.0%	7.02
FM	1	S9_144119131	9	144119131	$5.78X10^{-06}$	7.6%	9.06
	1	S6_150165479	6	150165479	$6.58X10^{-06}$	7.4%	6.05
	1	S7_13731608	7	13731608	$7.25X10^{-06}$	7.3%	7.01
	2	S5_217372319	5	217372319	$1.05X10^{-06}$	8.8%	5.09
	2	S2_222099831	2	222099831	$5.57X10^{-06}$	7.6%	2.08
	2	S1_16300644	1	16300644	$9.18X10^{-06}$	7.2%	1.02
	2	S1_65858162	1	65858162	$1.02X10^{-05}$	7.2%	1.04
	2	S7_8758861	7	8758861	$1.16X10^{-05}$	7.1%	7.01
	2	S3_218088466	3	218088466	$1.65X10^{-05}$	6.8%	3.09

associated with flowering time and reproductive processes. These genes were found to be involved in biological processes related to ASI (20 genes), FF (19 genes), and FM (54 genes) (Supplementary Table S5). The analysis also identified two co-functional networks that were found to be significantly enriched for genes related to single-organism reproductive behavior and the regulation of flower and reproductive development ($p<0.0005$). These networks identified four genes that were directly associated with the traits of FF and ASI (GRMZM2G114793 and GRMZM2G415007), and FM (GRMZM2G055520 and GRMZM2G161913) as shown in Figure 5.

The genes GRMZM2G114793 (bip1 - Binding protein homolog 1) and GRMZM2G415007 (bip2 - Binding protein homolog 2) were found to have orthologs in *Arabidopsis thaliana*, which encode BINDING PROTEIN 3.). The genes GRMZM2G055520 and GRMZM2G161913 have orthologs in *Arabidopsis thaliana* that encode EARLY FLOWERING 7 and EARLY FLOWERING 8, respectively. These genes are known to play a role in the control of flowering time in plants. Additionally, these four genes (GRMZM2G114793, GRMZM2G415007, GRMZM2G055520, and GRMZM2G161913) have an ontology associated with the stage of anthesis, or the beginning of flowering, in various cereal plants, including the silking stage in maize and the whole plant flowering stage.

3.4 Genomic prediction

The performance of four approaches (UTUE, UTME, MTUE, and MTME) for predicting flowering traits in tropical maize were compared using Bayesian and deep learning models. For this purpose, the approaches were evaluated in two scenarios: 1) selection of random training and validation datasets in each environment, and 2) prediction of Iguatemi season 2 using other environments as training datasets (Iguatemi season 1, DT1, Cambira and Sabaudia: DT2, and Cambira and Sabaudia + Iguatemi season 1: DT3).

3.4.1 Selection randomly into independent training and validation datasets in each environment (scenario I)

Predicting accuracy for uni-trait and multi-trait approaches in a single-environment (UTUE and MTUE):

The results of the study indicate that the multi-trait approach leads to higher prediction accuracy for all traits in each of the four environments evaluated, compared to the uni-trait approach. According to Table 2, prediction accuracies ranged from 0.11 to 0.73 for the uni-trait approach and from 0.16 to 0.73 for the multi-trait approach. The multi-trait approach, using the deep learning model, yielded the highest prediction accuracy (Table 2). On

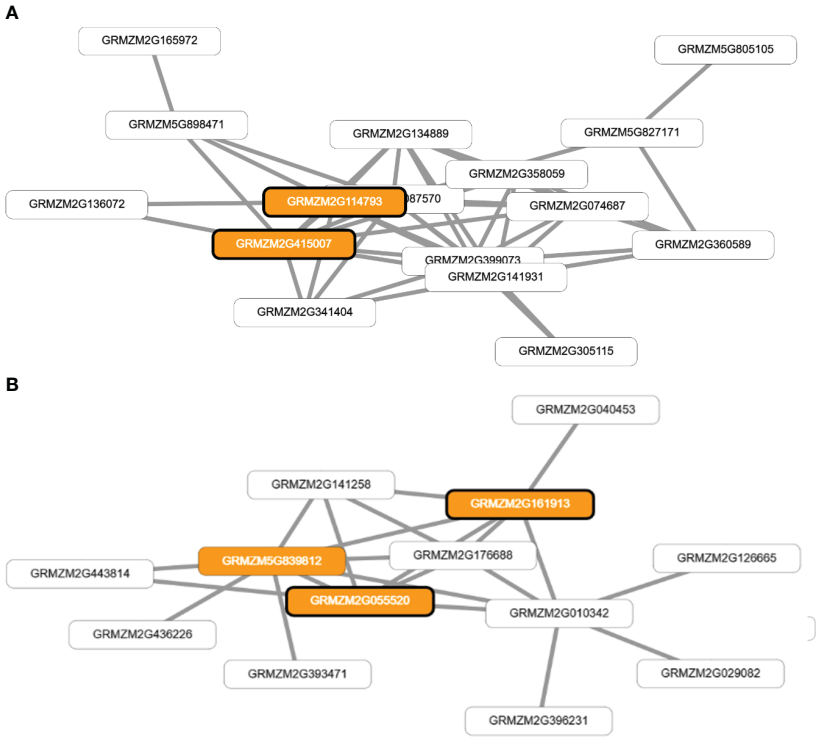


FIGURE 5
Visual representation of co-functional networks for flowering time traits in tropical maize. Panel **(A)** shows the network for anthesis-silking interval (ASI) and female flowering (FF) traits, while panel **(B)** displays the network for male flowering (FM). The networks were constructed using candidate genes identified through genome-wide association studies and prioritized using MaizeNet, a database of maize functional genomics. White boxes denote all the genes in the network, while orange boxes highlight genes that are associated with biological processes related to flowering time and reproduction, as identified by gene ontology (GO) annotations. The orange boxes with bold borders indicate genes identified by GWAS or through the prioritization analysis in MaizeNet.

average, multi-trait genomic selection models (MCMCglmm and deep learning) had higher (not significantly) prediction accuracy than uni-trait genomic selection models. Particularly, the largest improvement in prediction accuracy (26.6%) was observed when using the multi-trait approach with the MCMCglmm model, while the smallest improvement (2.2%) was observed when using the deep learning model. However, the highest prediction accuracy was obtained using the deep learning model for all traits in each of the environments, when UTUE and MTUE approaches were considered (Table 2). This suggests that the deep learning model is less sensitive to the use of uni- or multi-trait approaches. The highest (not significantly) prediction accuracies were obtained for the Cambira environment, while the lowest (not significantly) was obtained for the Iguatemi Season 2 environment, for all traits in both uni- and multi-trait approaches, and for both the MCMCglmm and deep learning models.

Predicting accuracies for uni-trait and multi-trait approaches in multi-environments (UTME and MTME):

In contrast to the analysis of a single environment, the prediction accuracy of the uni-trait model was found to be higher than that of the multi-trait model in most cases, as shown in Table 2. The prediction accuracies ranged from 0.14 (for ASI, in Iguatemi season 1, using BGGE and deep learning models) to 0.74 (for MF, in Cambira, using the BGGE model) and from 0.13 (for ASI, in Iguatemi season 1, using BMTME and deep learning models) to 0.72 (for MF, in Cambira, using the BMTME model) for the uni-trait and multi-trait approaches, respectively. Regardless of the GS model used, the multi-trait analysis showed lower (not significantly) prediction accuracies than the single-trait model, except in ASI Cambira (BMTME), ASI Sabaudia (DL) and MF Iguatemi season 2 (BMTME and DL). On average, the GS models in the single-trait analysis had 8.3% (BGGE-UTME over BMTME-MTME) and 4.5% (DL-UTME over DL-MTME) higher (not significantly) prediction accuracies than GS models in the multi-trait analysis. Similarly, in the single environment analysis, deep learning models had the highest prediction accuracy on average. Overall, the results indicate that the Uni-Trait-Multi-Environment and Multi-Trait-Uni-Environment approaches are more efficient for predicting flowering traits. Furthermore, deep learning consistently emerged as the most accurate model across all approaches, demonstrating superior performance across various traits and environments.

3.4.2 Prediction of flowering traits in Iguatemi season 2 (scenario II)

The prediction accuracy for scenario II was found to be generally higher than that of scenario I, as shown in Tables 2, 3. Table 3 presents the prediction accuracies for all flowering traits in the Iguatemi season 2 when the model was trained using three different datasets (DT1, DT2 and DT3) and four different approaches (Uni-Trait-Uni-Environment, Multi-Trait-Uni-Environment, Uni-Trait-Multi-Environment, and

TABLE 2 Predictive ability estimates for flowering time traits in a tropical maize panel across four environments (Cambira, Sabaudia, Iguatemi season 1, and Iguatemi season 2) using four different approaches: Uni-Trait-Uni-Environment (UTUE), Uni-Trait-Multi-Environment (UTME), Multi-Trait-Uni-Environment (MTUE), and Multi-Trait-Multi-Environment (MTME).

Environment	Trait	UTUE		UTME		MTUE		MTME	
		MCMCglmm	DL	BGGE	DL	MCMCglmm	DL	BMTME	DL
Cambira	FF	0.39e	0.62a	0.61ab	0.63a	0.45d	0.61ab	0.58c	0.59bc
	MF	0.42e	0.73ab	0.74a	0.73ab	0.49d	0.73bc	0.72bc	0.70c
	ASI	0.42bcd	0.5a	0.41d	0.44b	0.44bc	0.51a	0.41d	0.42cd
Sabaudia	FF	0.43c	0.49a	0.47ab	0.47ab	0.46abc	0.47ab	0.44bc	0.44abc
	MF	0.38e	0.58a	0.55abc	0.58ab	0.54bc	0.57abc	0.50d	0.56bc
	ASI	0.26c	0.31ab	0.30bc	0.30ab	0.29bc	0.32ab	0.28bc	0.33a
Iguatemi Season1	FF	0.36c	0.46ab	0.48a	0.46ab	0.45b	0.47a	0.46ab	0.42b
	MF	0.33d	0.55a	0.54a	0.51ab	0.45bc	0.54a	0.44c	0.48b
	ASI	0.12b	0.15ab	0.14ab	0.14ab	0.16ab	0.18a	0.13ab	0.13ab
Iguatemi Season2	FF	0.31d	0.42ab	0.43b	0.43b	0.42b	0.46a	0.33d	0.37c
	MF	0.29b	0.43a	0.31b	0.31b	0.29b	0.43a	0.31b	0.33b
	ASI	0.11c	0.27a	0.27a	0.25a	0.19b	0.27a	0.25a	0.24a

MCMCglmm, MCMC Generalised Linear Mixed Model; DL, Deep Learning; BGGE, Bayesian Genomic Genotype × Environment Interaction; BMTME, Bayesian Multi-Trait Multi-Environment. Statistical significance between different models is noted by lowercase letters. Different letters show the statistical significance at $p< 0.01$ according to the Tukey–Kramer test. The estimates are based on an average of 50 cross-validation cycles.

Multi-Trait-Multi-Environment). The prediction accuracies ranged from 0.20 to 0.66, 0.29 to 0.60, 0.24 to 0.68, and 0.13 to 0.60, respectively, for the four different approaches. It was found that the use of only season 1 of Iguatemi (DT1) performed the best among all scenarios. Additionally, the use of deep learning models was found to be more efficient (not significantly) for predicting flowering traits in tropical maize, with an improvement of 26.8% and 10.8% in the Uni-Trait and Multi-Trait approaches, respectively, when using DT1, and 2.5% and 1.6% in the Uni-Trait and Multi-Trait approaches, respectively, when using DT3.

4 Discussion

4.1 Genetic determinants of flowering traits in tropical maize

The flowering traits of crops are crucial for yield and seed quality (Helal et al., 2021). In this study, 31 significant SNP loci were identified that regulate flowering traits across two consecutive seasons. Of these, approximately 50% of significant SNPs were located on chromosomes 1, 2, and 3, which is consistent with

TABLE 3 Estimates of predictive ability for flowering time traits in a tropical maize panel for the second season of Iguatemi (validation dataset).

Scenario	Trait	UTUE		UTME		MTUE		MTME	
		MCMCglmm	DL	BGGE	DL	MCMCglmm	DL	BMTME	DL
DT1	FF	0.53	0.66	–	–	0.67	0.68	–	–
	MF	0.44	0.51	–	–	0.49	0.54	–	–
	ASI	0.20	0.28	–	–	0.24	0.29	–	–
DT2	FF	–	–	0.58	0.59	–	–	0.31	0.59
	MF	–	–	0.41	0.41	–	–	0.13	0.40
	ASI	–	–	0.31	0.29	–	–	0.20	0.34
DT3	FF	–	–	0.59	0.60	–	–	0.58	0.60
	MF	–	–	0.41	0.42	–	–	0.43	0.41
	ASI	–	–	0.31	0.32	–	–	0.34	0.36

MCMCglmm, MCMC Generalized Linear Mixed Model; DL, Deep Learning; BGGE, Bayesian Genomic Genotype × Environment Interaction; BMTME: Bayesian Multi-Trait Multi-Environment. DT1 (Iguatemi season 1), DT2 (Cambira and Sabaudia) and DT3 (Cambira, Sabaudia and the first season of Iguatemi) represent the training dataset. Four different approaches were used: Uni-Trait-Uni-Environment (UTUE), Uni-Trait-Multi-Environment (UTME), Multi-Trait-Uni-Environment (MTUE), and Multi-Trait-Multi-Environment (MTME) - indicates that the model was not run in this approach and scenario.

previous research that found over 33% of loci associated with flowering on these chromosomes (Li et al., 2016; Liu et al., 2019; Maldonado et al., 2019). Additionally, 9 SNPs (82%) associated with ASI were found on chromosomes 1, 3, and 4; for FF trait, 70% (7/10) of SNPs were found on chromosomes 2, 7, and 8; and for FM trait, 4 SNPs (44%) were distributed on chromosomes 1 and 7 (Supplementary Table S3). Previous studies have also identified significant SNPs associated with flowering traits in maize on similar chromosomes (Li et al., 2016; Liu et al., 2019; Maldonado et al., 2019; Shi et al., 2022), suggesting that these regions may contain genes that play a critical role in controlling flowering time variation in maize. The phenotypic variation explained by significant SNPs in this study ranged from 6.7 to 9.6% and was evenly distributed among traits, indicating that many significant SNPs of small effects contribute to genetic variation in flowering time in maize (Maldonado et al., 2019).

The study identified two potential pleiotropic loci that had an impact on both female and male flowering traits. The use of multivariate Bayesian regression (as suggested by Maldonado et al., 2019) allowed for the detection of pleiotropic genetic variants that are correlated with multiple traits by analyzing the Bayes factor and PPA. The PPA values of 0.99 and 0.74 for SNPs S6_150165479 and S5_217372319, respectively, provided strong evidence of the simultaneous association of these two loci with both FF and FM traits. Additionally, the high values of log10 (BF) (> 5.1) were considered to be strong evidence against the null hypothesis of no association and were higher than those found in previous association studies (Legarra et al., 2018). The correlation analysis results also showed a high and significant correlation between FF and FM, which supports the idea that FF and FM share similar loci. The study also found similarities to previous research by Li et al. (2016) who identified two pleiotropic significant SNPs located in the same bin (6.05) of loci S6_150165479, indicating that this region affects both female and male flowering time. These discoveries of pleiotropic significant SNPs could aid in understanding the molecular mechanisms of flowering time in maize.

GWAS is a powerful tool for identifying genetic variants associated with specific traits in maize. Studies such as those by Xiao et al. (2016); Coan et al. (2018); Maldonado et al. (2019), and Shi et al. (2022) have used GWAS to identify key genetic variants underlying phenotypic variation in several maize traits. Additionally, Wallace et al. (2014) found that the majority of the variance in maize can be explained by within-gene and gene-proximal SNPs (at about 1–5 kb). By using high-resolution GWAS, it may be possible to identify loci that significantly affect maize flowering time within candidate genes or in proximity to them. Therefore, GWAS approaches can be a useful tool for understanding the genetic basis of flowering time in maize and for identifying potential targets for crop improvement.

The association analysis identified several markers associated with flowering traits in maize, which explain up to 9.6% of phenotypic variation individually, and between 67 and 86% of the trait phenotypic variation considering all significant markers. This result is consistent with previous studies on traits related to flowering time in maize (Salvi et al., 2009; Liu et al., 2019; Maldonado et al., 2019).Moreover, it is worth noting that the LD pattern exhibits a rapid decline within a 2.7 kb range, which aligns with the findings reported by Coan et al. (2018) and Maldonado et al. (2019). This LD pattern indicated that candidate genes should be located within a 2.7 kb region upstream and downstream of significant SNPs. The gene-prioritization and co-functional network approach found that four genes were significantly associated with the stage at flowers open, anthesis and silking in some cereal plants such as maize. In this regard, hundreds of genes in plants have been extensively studied in *Arabidopsis*. In this study, ortholog genes for BINDING PROTEIN 3 (which control pollen germination and pollen tube elongation; Sato and Maeshima, 2019), orthologs associated with the stage of anthesis or the beginning of flowering (particularly important in the sporophyte reproductive stage; Xiang et al., 2011), and EARLY FLOWERING genes (which play a crucial role in determining when a plant flower; Li et al., 2016; Li et al., 2019) have been identified. Particularly, two orthologs of EARLY FLOWERING genes, ELF7 and ELF8, were identified as candidate genes controlling flowering time in maize using gene-prioritization and subnetwork analysis of the MaizeNet database (Lee et al., 2019). These genes have been shown to cause rapid flowering in various situations where flowering would otherwise be delayed (He et al., 2004; Li et al., 2016). Additionally, ELF7 and ELF8 are known to regulate the expression of genes in the FLOWERING LOCUS C clade, which includes repressors such as MAF2 and FLM that play a role in multiple flowering pathways (He et al., 2004). The SNPs and candidate genes associated with flowering time phenotypes identified in this study can be integrated into molecular marker-assisted breeding programs and provide valuable genetic resources for future maize breeding efforts.

4.2 Multi-trait and multi-environment genomic prediction for flowering traits in maize

Genomic selection is a powerful strategy that has been proven to significantly improve the efficiency of breeding programs by increasing genetic gain and reducing selection time (Bhat et al., 2016). The goal of GS is to construct accurate prediction models using training populations that consist of individuals with both genotypic and phenotypic data. In practice, plant breeders often collect data for multiple traits in different environments and over multiple years. Studies have shown that prediction approaches based on Multi-Trait and Multi-Environments (MT-ME) are more accurate than Uni-Trait and Uni-Environment (UT-UE) approaches because they allow for the prediction of multiple traits simultaneously, which reduces the number of locations needed for subsequent selection trials (Tolhurst et al., 2019; Larkin et al., 2021; Sandhu et al., 2022). Despite the benefits of using MT-ME approaches, few GS studies have adopted them due to the complexity of the models (Cuevas et al., 2017). Therefore, in this study, different models based on MT-ME approaches were evaluated and compared with UT-UE approaches to predict flowering traits in inbred lines of tropical maize.

In the scenario I, when considering selection randomly into independent training and validation datasets in each environment,

the multi-trait approach performed 14.4% superior to the Uni-Trait approach for the Uni-environment, while in the Multi-Environment approach, the Uni-Trait approach performed 6.4% superior to the Multi-Trait approach. Notably, regardless of the approach, the Deep Learning model showed a higher prediction accuracy (Table 2). Additionally, the Deep Learning model was significantly superior to the MCMCglmm (Uni-Trait-Uni-Environment and Multi-Trait-Uni-Environment) and BMTME (Multi-Trait-Multi-Environment) models. These results may be due to the ability of the Deep Learning model to automatically capture complex interactions in its hidden layers without the need to specify the covariates corresponding to interactions between traits or environments in the predictor, as previously noted by Montesinos-López et al. (2018). It is worth noting that similar results have been observed by Montesinos-López et al. (2018) where the Deep Learning model performed superiorly to other models when the genotype-environment interaction (Uni-Environment) is not considered, but its advantages diminished when the genotype-environment component is included in the model, which is consistent with the findings of this study.

In scenario II, when predicting the second season of Iguatemi, utilizing information from the first season of Iguatemi (DT1) was found to be more accurate than utilizing information from other environments (DT2 and DT3). This may be due to the high correlation observed between the first and second seasons of Iguatemi for traits such as FF (r = 0.68), MF (r = 0.53), and anthesis-silking interval (ASI: r = 0.30), compared to the correlation between these traits in other environments. Furthermore, the results of this scenario differed from those of scenario I, as the prediction accuracy for the FF trait was found to be superior to that of the MF trait. This may be due to the high correlation observed among all environments and traits for the FF trait (as shown in Figures 2, 3), as previously reported by Sandhu et al. (2022) and Montesinos-López et al. (2016), who mention that a high correlation between traits improves prediction accuracies and highlights the importance of using multi-trait models. Additionally, for the ASI trait, which has a low correlation among traits and environments, as well as a low heritability (h^2 = 0.29), the Deep Learning model in the DT3 (Multi-Trait-Multi-Environment) approach was found to be more effective than models in the DT1 (Uni-Trait-Uni-Environment and Multi-Trait-Uni-Environment) approach. This suggests that Multi-Trait-Multi-Environment approaches may be useful for increasing predictions for primary traits with low heritability when a secondary trait is highly correlated and has high heritability (as reported by Sandhu et al., 2022). As noted by Cui et al. (2020), heritability can vary depending on the genetic architecture of traits, with traits such as flowering date being controlled by several major genetic loci that have high heritabilities. This study found that the flowering traits had moderate to high heritabilities (FF: 0.72, MF: 0.66, and ASI: 0.29) (as reported by Cui et al., 2020; Maldonado et al., 2020). As expected, the prediction accuracy was moderate to high (as reported by Zhang et al., 2017), with higher prediction accuracies observed for traits with higher heritability compared to those with lower heritability. Similar results have been observed in previous studies, with high positive correlations between heritability and prediction accuracy values (as reported by Nyine et al., 2017; Cui et al., 2020; Kaler et al., 2022). Notably, the Deep Learning model showed higher prediction accuracy compared to other models, regardless of the heritability of the trait. This is in line with the findings of Alves et al. (2020) who found that artificial neural network models had a higher prediction accuracy compared with GBLUP for traits with moderate heritability, indicating that neural network models may be a promising alternative tool for genomic prediction, independent of the contribution of genetic effects (as reported by Maldonado et al., 2020).

In all scenarios, the use of Deep Learning models resulted in higher prediction accuracy compared to other models for all traits (except BGGE in UTME, since it had similar predictions). This suggests that the Deep Learning model is less sensitive to random variations among seasons and correlations between traits and that it does not require the consideration of "genotype x environment" interactions and prior information on the covariance matrices of traits (genetic and residual) for training and constructing the predictive model (Montesinos-López et al., 2018). In this regard, Maldonado et al. (2020) highlighted that machine learning-based GP models can treat response variables as an implicit function of input variables (e.g., environmental components) through non-linear and highly complex functions, which implies that these models can effectively increase prediction accuracy without the need to pre-specify interaction terms.

In this study, it was shown that Deep Learning models based on Uni- or Multi-Trait and Uni- or Multi-Environment approaches outperformed Bayesian Genomic Selection models (MCMCglmm and BMTME). It should be noted that BGGE achieved the same level of prediction accuracy as DL in UTME, however, the computational time required for BGGE was approximately three times longer than that of DL (data not shown). Similar results were observed by Maldonado et al. (2020), which indicated that DL models require significantly less computational time (approximately 16 times less) compared to traditional Bayesian models. The superiority of Deep Learning models in GS over traditional mixed model-based approaches has been previously reported in the literature by Sandhu et al. (2021), Zingaretti et al. (2020), Montesinos-López et al. (2018), and Maldonado et al. (2020). According to Sandhu et al. (2022), Deep Learning models are highly flexible in understanding the complex interactions present in datasets, and they can infer trends present in datasets better than traditional models. The results of this study confirm the importance of Deep Learning models for increasing prediction accuracy in GS, which holds promise for accelerating crop breeding progress.

5 Conclusion

In conclusion, this study highlights the effectiveness of deep learning models in genomic selection studies for predicting complex flowering-related traits in tropical maize. Deep learning models outperformed other models (except for BGGE in UTME where similar predictions were observed) indicating their superior accuracy across all traits and scenarios. This suggests that Multi-Traits deep learning models are less affected by low or negative correlations among traits. Moreover, these models have the

advantage of learning patterns directly from the data without relying on prior assumptions, making them an attractive alternative to traditional Multi-Trait and Multi-Environment based models. Among the deep learning models, the MTUE model consistently demonstrated the highest prediction accuracies on average. Therefore, it is recommended to use this model in breeding programs, especially for predicting traits that are challenging or expensive to phenotype, or those with low levels of correlation. Additionally, deep learning models should be incorporated into the toolkit of plant breeders to accelerate crop breeding progress and improve genetic gain for quantitative traits. On the other hand, this study identified several loci in genomic regions associated with flowering time in tropical maize, which have variable contributions to phenotypic expression. These findings can be utilized in marker-assisted selection programs, where the loci identified can be target to improve breeding outcomes. Additionally, through the co-functional network approach (post-GWAS), orthologs of EARLY FLOWERING genes were identified, which offer potential targets for genome editing programs focused on improving flowering traits. These discoveries provide valuable insights into the genetic architecture and underlying mechanisms of flowering-related traits in tropical maize, which can be incorporated into breeding programs for further advancements.

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

FM-P, CM, and CS conceived the research plans. LH, CAM, and RU performed the data curation. CM and FM-P analyzed the genomic data and wrote the first draft of the manuscript. CAM, LH, and CS supervised the field experiments. FM-P, CAM, RU, and LH reviewed and edited the final version of manuscript. All authors reviewed and approved the paper for publication. All authors contributed to the article and approved the submitted version.

Funding

This research was supported by CNPq and CAPES, Brazil.

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1153040/full#supplementary-material

References

Alves, M. L., Belo, M., Carbas, B., Brites, C., Paulo, M., Mendes-Moreira, P., et al. (2018). Long-term on-farm participatory maize breeding by stratified mass selection retains molecular diversity while improving agronomic performance. *Evol. Applications* 11, 254–270. doi: 10.1111/eva.12549

Alves, A. A. C., da Costa, R. M., Bresolin, T., Fernandes Júnior, G. A., Espigolan, R., Ribeiro, A. M. F., et al. (2020). Genome-wide prediction for complex traits under the presence of dominance effects in simulated populations using GBLUP and machine learning methods. *J. Anim. Sci.* 98, skaa179. doi: 10.1093/jas/skaa179

Bhat, J. A., Ali, S., Salgotra, R. K., Mir, Z. A., Dutta, S., Jadon, V., et al. (2016). Genomic selection in the era of next generation sequencing for complex traits in plant breeding. *Front. Genet.* 7. doi: 10.3389/fgene.2016.00221

Birnbaum, K. D., and Roberts, J. K. (2019). Identification of QTLs for flowering time in a panel of maize inbred lines. *Theor. Appl. Genet.* 132 (7), 1835–1846.

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

Breseghello, F., and Sorrells, M. E. (2006). Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. *Genetics* 172, 1165–1177. doi: 10.1534/genetics.105.044586

Chan, D. M., Rao, R., Huang, F., and Canny, J. F. (2018). "T-SNE-CUDA: GPU-accelerated T-SNE and its applications to modern data," in *2018 30th International Symposium on Computer Architecture and High Performance Computing (SBAC-PAD)*, Lyon, France. 2018, 330–338. doi: 10.1109/CAHPC.2018.8645912

Chen, D. H., and Ronald, P. C. (1999). A rapid DNA minipreparation method suitable for AFLP and other PCR applications. *Plant Mol. Biol. Rep.* 17, 53–57. doi: 10.1023/A:1007585532036

Coan, M., Senhorinho, H. J., Pinto, R. J., Scapim, C. A., Tessmann, D. J., Williams, W. P., et al. (2018). Genome-wide association study of resistance to ear rot by Fusarium verticillioides in a tropical field maize and popcorn core collection. *Crop Sci.* 58, 564–578. doi: 10.2135/cropsci2017.05.0322

Costa-Neto, G., Fritsche-Neto, R., and Crossa, J. (2020). Nonlinear kernels, dominance, and envirotyping data increase the accuracy of genome-based prediction in multi-environment trials. *Heredity (Edinb).* 126, 92–106. doi: 10.1038/s41437-020-00353-1

Crossa, J., Martini, J. W. R., Gianola, D., Perez-Rodriguez, P., Jarquin, D., Juliana, P., et al. (2019). Deep kernel and deep learning for genome-based prediction of single traits in multienvironment breeding trials. *Front. Genet.* 10. doi: 10.3389/fgene.2019.01168

Cuevas, J., Crossa, J., Montesinos-López, O. A., Burgueño, J., Pérez-Rodríguez, P., and de los Campos, G. (2017). Bayesian genomic prediction with genotype ×

Environment interaction kernel models. *G3 Genes Genomes Genet.* 7, 41–53. doi: 10.1534/G3.116.035584

Cui, Z., Dong, H., Zhang, A., Ruan, Y., He, Y., and Zhang, Z. (2020). Assessment of the potential for genomic selection to improve husk traits in maize. *G3: Genes Genomes Genet.* 10, 3741–3749. doi: 10.1534/G3.120.401600

Domínguez-Hernández, E., Gaytán-Martínez, M., Gutiérrez-Uribe, J. A., and Domínguez-Hernández, M. E. (2022). The nutraceutical value of maize (*Zea mays* L.) landraces and the determinants of its variability. *A review. J. Cereal Sci.* 103, 103399. doi: 10.1016/j.jcs.2021.103399

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

Endelman, J. B., and Jannink, J. L. (2012). Shrinkage estimation of the realized relationship matrix. *G3* 2, 1405–1413. doi: 10.1534/g3.112.004259

Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* 14, 2611–2620. doi: 10.1111/j.1365-294X.2005.02553.x

Gao, H., Williamson, S., and Bustamante, C. D. (2007). A Markov chain Monte Carlo approach for joint inference of population structure and inbreeding rates from multilocus genotype data. *Genetics* 176, 1635–1651. doi: 10.1534/genetics.107.072371

Gedil, M., and Menkir, A. (2019). An integrated molecular and conventional breeding scheme for enhancing genetic gain in maize in Africa. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.01430

Gill, H. S., Halder, J., Zhang, J., Brar, N. K., Rai, T. S., Hall, C., et al. (2021). Multi-trait multi-environment genomic prediction of agronomic traits in advanced breeding lines of winter wheat. *Front. Plant Sci.* 0. doi: 10.3389/FPLS.2021.709545

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, e90346. doi: 10.1371/journal.pone.0090346

Granato, I., Cuevas, J., Luna-Vazquez, F., Crossa, J., Montesinos-Lopez, O., Burgueno, J., et al. (2018). BGGE: a new package for genomic-enabled prediction incorporating genotype x environment interaction models. *G3 (Bethesda)* 8, 3039–3047. doi: 10.1534/g3.118.200435

Grote, U., Fasse, A., Nguyen, T. T., and Erenstein, O. (2021). Food security and the dynamics of wheat and maize value chains in Africa and Asia. *Front. Sustain. Food Syst.* 4. doi: 10.3389/fsufs.2020.617009

Hadfield, J. D., and Nakagawa, S. (2010). General quantitative genetic methods for comparative biology: phylogenies, taxonomies and multi-trait models for continuous and categorical characters. *J. Evol. Biol.* 23, 494–508. doi: 10.1111/j.1420-9101.2009.01915.x

He, Y., Doyle, M. R., and Amasino, R. M. (2004). PAF1-complex-mediated histone methylation of FLOWERING LOCUS C chromatin is required for the vernalization-responsive, winter-annual habit in Arabidopsis. *Genes Dev.* 18, 2774–2784. doi: 10.1101/gad.1244504

Helal, M. M., Gill, R. A., Tang, M., Yang, L., Hu, M., Yang, L., et al. (2021). SNP and haplotype-based GWAS of flowering-related traits in Brassica napus. *Plants* 10, 2475. doi: 10.3390/plants10112475

Hirohata, A., Yamatsuta, Y., Ogawa, K., Kubota, A., Suzuki, T., Shimizu, H., et al. (2022). Sulfanilamide regulates flowering time through expression of the circadian clock gene LUX. *Plant Cell Physiol.* 63 (5), 649–657. doi: 10.1093/pcp/pcac027

Hu, H., Meng, Y., Liu, W., Chen, S., and Runcie, D. E. (2022). Multi-trait genomic prediction improves accuracy of selection among doubled haploid lines in maize. *Int. J. Mol. Sci.* 23 (23), 14558. doi: 10.3390/ijms232314558

Kaler, A. S., Purcell, L. C., Beissinger, T., and Gillman, J. D. (2022). Genomic prediction models for traits differing in heritability for soybean, rice, and maize. *BMC Plant Biol.* 22 (1), 1–11. doi: 10.1186/s12870-022-03479-y

Larkin, D. L., Mason, R. E., Moon, D. E., Holder, A. L., Ward, B. P., and Brown-Guedira, G. (2021). Predicting fusarium Head Blight Resistance for Advanced Trials in a Soft Red winter Wheat Breeding Program with Genomic Selection. *Front. Plant Sci.* 12. doi: 10.3389/FPLS.2021.715314

Lee, T., Lee, S., Yang, S., and Lee, I. (2019). MaizeNet: a co-functional network for network-assisted systems genetics in Zea mays. *Plant J.* 99, 571–582. doi: 10.1111/tpj.14341

Legarra, A., Ricard, A., and Varona, L. (2018). GWAS by GBLUP: single and multimarker EMMAX and bayes factors, with an example in detection of a major gene for horse gait. *G3* 8, 2301–2308. doi: 10.1534/g3.118.200336

Leng, P., Khan, S. U., Zhang, D., Zhou, G., Zhang, X., Zheng, Y., et al. (2022). Linkage mapping reveals QTL for flowering time-related traits under multiple abiotic stress conditions in maize. *Int. J. Mol. Sci.* 23 (15), 8410. doi: 10.3390/ijms23158410

Li, Y.-X., Li, C., Bradbury, P. J., Liu, X., Lu, F., Romay, C. M., et al. (2016). Identification of genetic variants associated with maize flowering time using an extremely large multi-genetic background population. *Plant J.* 86, 391–402. doi: 10.1111/tpj.13174

Li, Y., Yang, J., Shang, X. D., Lv, W. Z., Xia, C. C., Wang, C., et al. (2019). SKIP regulates environmental fitness and floral transition by forming two distinct complexes in Arabidopsis. *New Phytol.* 224, 321–335. doi: 10.1111/nph.15990

Liu, Y., Hu, G., Zhang, A., Loladze, A., Hu, Y., Wang, H., et al. (2021). Genome-wide association study and genomic prediction of Fusarium ear rot resistance in tropical maize germplasm. *Crop J.* 9, 325–341. doi: 10.1016/j.cj.2020.08.008

Liu, S., Zenda, T., Wang, X., Liu, G., Jin, H., Yang, Y., et al. (2019). Comprehensive meta-analysis of maize QTLs associated with grain yield, flowering date and plant height under drought conditions. *J. Agric. Sci.* 11, 1–19. doi: 10.5539/jas.v11n8p1

López-Cortés, X. A., Matamala, F., Maldonado, C., Mora-Poblete, F., and Scapim, C. A. (2020). A deep learning approach to population structure inference in inbred lines of maize. *Front. Genet.* 11, 543459. doi: 10.3389/fgene.2020.543459

Ma, J., and Cao, Y. (2021). Genetic dissection of grain yield of maize and yield-related traits through association mapping and genomic prediction. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.690059

Maldonado, C., Mora, F., Bertagna, F. A. B., Kuki, M. C., and Scapim, C. A. (2019). SNP-and haplotype-based GWAS of flowering-related traits in maize with network-assisted gene prioritization. *Agronomy* 9, 725. doi: 10.3390/agronomy9110725

Maldonado, C., Mora-Poblete, F., Contreras-Soto, R. I., Ahmar, S., Chen, J.-T., do Amaral Júnior, A. T., et al. (2020). Genome-wide prediction of complex traits in two outcrossing plant species through deep learning and bayesian regularized neural network. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.593897

Marchini, J., and Band, G. (2016) *SNPTEST*. Available at: https://mathgen.stats.ox.ac.uk/genetics_software/snptest/snptest.html.

Mathew, B., Holand, A. M., Koistinen, P., Léon, J., and Sillanpää, M. J. (2016). Reparametrization-based estimation of genetic parameters in multi-trait animal model using Integrated Nested Laplace Approximation. *Theor. Appl. Genet.* 129, 215–225. doi: 10.1007/s00122-015-2622-x

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

Montesinos-López, O. A., Montesinos-López, A., Crossa, J., Gianola, D., Hernández-Suárez, C. M., and Martín-Vallejo, J. (2018). Multi-trait, multi-environment deep learning modeling for genomic-enabled prediction of plant traits. *G3 (Bethesda)* 8, 3829–3840. doi: 10.1534/g3.118.200728

Montesinos-López, O. A., Montesinos-López, A., Crossa, J., Toledo, F. H., Pérez-Hernández, O., Eskridge, K. M., et al. (2016). A genomic Bayesian multi-trait and multi-environment model. *G3 Genes|Genomes|Genetics* 6, 2725–2744. doi: 10.1534/g3.116.032359

Montesinos-López, O. A., Montesinos-López, A., Tuberosa, R., Maccaferri, M., Sciara, G., Ammar, K., et al. (2019). Multi-trait, multi-environment genomic prediction of durum wheat with genomic best linear unbiased predictor and deep learning methods. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.01311

Nyine, M., Uwimana, B., Swennen, R., Batte, M., Brown, A., Christelová, P., et al. (2017). Trait variation and genetic diversity in a banana genomic selection training population. *PloS One* 12, e0178734. doi: 10.1371/journal.pone.0178734

Parent, B., Leclere, M., Lacube, S., Semenov, M. A., Welcker, C., Martre, P., et al. (2018). Maize yields over Europe may increase in spite of climate change, with an appropriate use of the genetic variability of flowering time. *Proc. Natl. Acad. Sci. U.S.A.* 115, 10642–10647. doi: 10.1073/pnas.1720716115

Romero, J. A., Willcox, M., Burgueño, J., Romay, C., Swarts, K., Trachsel, S., et al. (2017). A study of allelic diversity underlying flowering-time adaptation in maize landraces. *Nat. Genet.* 49, 476–480. doi: 10.1038/ng.3784

Salvi, S., Castelletti, S., and Tuberosa, R. (2009). An updated consensus map for flowering time QTLs in maize. *Maydica* 54, 501–512.

Sandhu, K. S., Patil, S. S., Aoun, M., and Carter, A. H. (2022). Multi-trait multienvironment genomic prediction for end-use quality traits in winter wheat. *Front. Genet.* 13. doi: 10.3389/fgene.2022.831020

Sandhu, K., Patil, S. S., Pumphrey, M., and Carter, A. (2021). Multitrait machine-and deep-learning models for genomic selection using spectral information in a wheat breeding program. *Plant Genome* 14, e20119. doi: 10.1002/TPG2.20119

Sato, R., and Maeshima, M. (2019). The ER-localized aquaporin SIP2; 1 is involved in pollen germination and pollen tube elongation in Arabidopsis thaliana. *Plant Mol. Biol.* 100, 335–349. doi: 10.1007/s11103-019-00865-3

Shi, J., Wang, Y., Wang, C., Wang, L., Zeng, W., Han, G., et al. (2022). Linkage mapping combined with GWAS revealed the genetic structural relationship and candidate genes of maize flowering time-related traits. *BMC Plant Biol.* 22 (1), 1–13. doi: 10.1186/s12870-022-03711-9

Stephens, M., and Balding, D. J. (2009). Bayesian statistical methods for genetic association studies. *Nat. Rev. Genet.* 10, 681–690. doi: 10.1038/nrg2615

Strable, J., and Scanlon, M. J. (2009). Maize (Zea mays): A model organism for basic and applied research in plant biology. *Cold Spring Harb. Protoc.* 2009, emo132. doi: 10.1101/pdb.emo132

Team R. C (2013). *R: A Language and Environment for Statistical Computing* (Davis, CA: Team, R.C).

Tolhurst, D. J., Mathews, K. L., Smith, A. B., and Cullis, B. R. (2019). Genomic selection in multi-environment plant breeding trials using a factor analytic linear mixed model. *J. Anim. Breed. Genet.* 136, 279–300. doi: 10.1111/JBG.12404

Torres, L. G., Rodrigues, M. C., Lima, N. L., Trindade, T. F. H., Silva, F. F., and Azevedo, C. F. (2018). Multi-trait multi-environment Bayesian model reveals G x E

interaction for nitrogen use efficiency components in tropical maize. *PloS One* 13, e0199492. doi: 10.1371/journal.pone.0199492

VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91, 4414–4423. doi: 10.3168/jds.2007-0980

Vinayan, M. T., Seetharam, K., Babu, R., Zaidi, P. H., Blummel, M., and Nair, S. K. (2021). Genome wide association study and genomic prediction for stover quality traits in tropical maize (*Zea mays* L.). *Sci. Rep.* 11, 686. doi: 10.1038/s41598-020-80118-2

Wallace, J. G., Bradbury, P. J., Zhang, N. Y., Gibon, Y., Stitt, M., and Buckler, E. S. (2014). Association mapping across numerous traits reveals patterns of functional variation in maize. *PloS Genet.* 10, 1–10. doi: 10.1371/journal.pgen.1004845

Xiang, L., Le Roy, K., Bolouri-Moghaddam, M. R., Vanhaecke, M., Lammens, W., Rolland, F., et al. (2011). Exploring the neutral invertase-oxidative stress defence connection in Arabidopsis thaliana. *J. Exp. Bot.* 62, 3849–3862. doi: 10.1093/jxb/err069

Xiao, Y., Tong, H., Yang, X., Xu, S., Pan, Q., Feng, Q., et al. (2016). Genome-wide dissection of the maize ear genetic architecture using multiple populations. *New Phytol.* 210, 1095–1106. doi: 10.1111/nph.13814

Zhang, A., Wang, H., Beyene, Y., Semagn, K., Liu, Y., Cao, S., et al. (2017). 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. doi: 10.3389/fpls.2018.1916

Zhou, G., Zhu, Q., Mao, Y., Chen, G., Xue, L., Lu, H., et al. (2021). Multi-locus genome-wide association study and genomic selection of kernel moisture content at the harvest stage in maize. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.697688

Zingaretti, L. M., Gezan, S. A., Ferrão, L. F. V., Osorio, L. F., Monfort, A., Muñoz, P. R., et al. (2020). Exploring deep learning for complex trait genomic prediction in polyploid outcrossing species. *Front. Plant Sci.* 11, 25. doi: 10.3389/fpls.2020.00025

frontiers | Frontiers in Plant Science

TYPE Review
PUBLISHED 19 September 2023
DOI 10.3389/fpls.2023.1249230

OPEN ACCESS

EDITED BY
Baohong Zhang,
East Carolina University, United States

REVIEWED BY
Phetole Mangena,
University of Limpopo, South Africa
Christell Van Der Vyver,
Stellenbosch University, South Africa

*CORRESPONDENCE
Dharam P. Chaudhary
chaudharydp@gmail.com

RECEIVED 28 June 2023
ACCEPTED 01 September 2023
PUBLISHED 19 September 2023

CITATION
Devi V, Bhushan B, Gupta M, Sethi M, Kaur C, Singh A, Singh V, Kumar R, Rakshit S and Chaudhary DP (2023) Genetic and molecular understanding for the development of methionine-rich maize: a holistic approach. *Front. Plant Sci.* 14:1249230. doi: 10.3389/fpls.2023.1249230

Genetic and molecular understanding for the development of methionine-rich maize: a holistic approach

Veena Devi[1], Bharat Bhushan[1], Mamta Gupta[2], Mehak Sethi[1], Charanjeet Kaur[3], Alla Singh[2], Vishal Singh[4], Ramesh Kumar[4], Sujay Rakshit[4] and Dharam P. Chaudhary[1]*

[1]Division of Biochemistry, Indian Institute of Maize Research, Ludhiana, Punjab, India, [2]Division of Biotechnology, Indian Institute of Maize Research, Ludhiana, Punjab, India, [3]Department of Biochemistry, Punjab Agricultural University, Ludhiana, Punjab, India, [4]Division of Plant Breeding, Indian Institute of Maize Research, Ludhiana, Punjab, India

Maize (*Zea mays*) is the most important coarse cereal utilized as a major energy source for animal feed and humans. However, maize grains are deficient in methionine, an essential amino acid required for proper growth and development. Synthetic methionine has been used in animal feed, which is costlier and leads to adverse health effects on end-users. Bio-fortification of maize for methionine is, therefore, the most sustainable and environmental friendly approach. The zein proteins are responsible for methionine deposition in the form of δ-zein, which are major seed storage proteins of maize kernel. The present review summarizes various aspects of methionine including its importance and requirement for different subjects, its role in animal growth and performance, regulation of methionine content in maize and its utilization in human food. This review gives insight into improvement strategies including the selection of natural high-methionine mutants, molecular modulation of maize seed storage proteins and target key enzymes for sulphur metabolism and its flux towards the methionine synthesis, expression of synthetic genes, modifying gene codon and promoters employing genetic engineering approaches to enhance its expression. The compiled information on methionine and essential amino acids linked Quantitative Trait Loci in maize and orthologs cereals will give insight into the hotspot-linked genomic regions across the diverse range of maize germplasm through meta-QTL studies. The detailed information about candidate genes will provide the opportunity to target specific regions for gene editing to enhance methionine content in maize. Overall, this review will be helpful for researchers to design appropriate strategies to develop high-methionine maize.

KEYWORDS

maize, methionine, δ-zein, poultry feed, QTLs, sulphur metabolism

1 Introduction

Maize is a globally important crop and among cereals occupies third place after rice (*Oryza sativa*) and wheat (*Triticum aestivum*). It is also called the "Queen of Cereals" because of its high genetic yield potential among cereals. The worldwide production of maize was around 1123.07 million metric tons (M MT) in 2020-2021 (Food and Agriculture Organization, 2022). The crop has tremendous genetic variability, which enables it to thrive in tropical, subtropical, and temperate climates. The USA, China, Brazil, European Union, Argentina, India, Ukraine, Mexico, South Africa and Russia are the top ten maize-producing countries. In India, maize is grown throughout the year. It is predominantly a *kharif* (rainy season) crop. Total maize production accounts for ~10% of total food grain production in the country. India contributes around 2.80% of maize production with a quantum of 31.51 M MT in 2020-2021 (Food and Agriculture Organization, 2022). Due to much lower water requirements under changing climate with dwindling water resources and higher CO_2 levels in the atmosphere, maize is going to play the most important role in world agriculture. Along with this, being a C_4 plant, it has higher photosynthetic efficiency as compared to C_3 plants.

Maize is consumed as a staple food in Africa, South America and some parts of Asia. In India, about 60% (18.91 M MT) of maize produced is used for livestock and poultry feed, 20% (6.30 M MT) as food, the rest is used as fuel (3.0 M MT) and for industrial purposes (3.2 M MT) (CROPS, 2018). The rapid expansion of the Indian population is changing the consumption patterns, especially of the urban population (Zhou and Staatz, 2016). The increased demand of poultry meat has resulted in the increased production of the feed crops such as maize and soybean (*Glycine max*).

Structurally, the maize kernel consists of an embryo, a much larger endosperm and a pericarp (Wu and Messing, 2014). Nutritionally, maize endosperm contains ~90% starch and 10% protein (Gibbon and Larkins, 2005). Endosperm storage proteins are classified as albumins (3%), globulins (3%), prolamins (also known as zein) (60%) and glutelins (34%). Maize grain protein is nutritionally imbalanced due to deficiency of essential amino acids (methionine [Met], lysine [Lys], tryptophan [Trp] and threonine [Thr]) in the prolamin fraction (Darrigues et al., 2005). As maize is used for feed, so to balance the nutrition in animal feed mixture, corn is usually supplemented with legumes, but Met remains a limiting amino acid in such feed mixtures as all legumes cannot supplement the deficiency of Met (Scott et al., 2004). The level of Met, Lys and Trp is more important than the total protein content in an animal feed mix. So, nutritionally balanced maize is important, considering the fact that the majority of the maize produced is consumed as food and feed.

Plants, unlike animals, synthesize Met *de novo* and are thus a dietary source of this nutrient for animals. However, supplementation of feed with synthetic amino acids and mineral mixture is a common practice in the feed sector. Thus, synthetic amino acids have become a multimillion-dollar industry (Lai and Messing, 2002). The total worldwide Met market was 685–700 million tons (MT) in 2017, which increased by 27% in 2018 (Xiang et al., 2018). Additionally, supplementation of essential amino acids increases the cost of feed, particularly in the case of Trp and Met for which inexpensive supplements are lacking (Darrigues et al., 2005). It was reported that increasing the content of Met in the diet significantly increases the weight of chicks (Panda et al., 2010). So, the maize varieties with enhanced levels of Met along with Trp and Lys, might have great economic potential in developing countries like India, where the feed sector is growing substantially. These Met bio-fortified maize cultivars will be able to replace the use of expensive synthetic Met. Keeping this in view, this article focused on the targeted strategies to enrich the maize germplasm with higher Met content.

1.1 Biochemistry and regulation of zein proteins

1.1.1 Biochemistry of zeins

Storage proteins in seeds act as a reservoir which is utilized during early seedling growth. All fractions of maize storage proteins, except prolamins (zeins) are balanced in their amino acid composition. The high proportion of zeins in the endosperm is the primary reason for the poor protein quality of maize (Vasal, 2000) as zeins proteins are deficient in essential amino acids such as Lys and Trp, whereas non-zeins are rich in these amino acids. In maize, among the zein proteins, β- and δ-zein proteins are rich in Met residues. However, these fractions do not accumulate at sufficient levels to balance Met content.

Various mutant maize varieties are available, like *o2* mutant maize lines, as they have high Lys content, but these lines failed to increase the level of Met, rather somewhat decreased its level (Mertz et al., 1964; Phillips and Mcclure, 1985; Wu et al., 2012). According to the maize protein database, 8% of the maize kernel proteins have Met content above 4%, while about 57% of the maize kernel proteins have Lys content above 4% (Wu et al., 2012). Therefore, for the rebalancing of maize seed storage protein composition, allele-specific up-regulation and gene-editing to enhance Met content would be suitable strategies.

Zeins have been categorized based on differences in solubility, molecular weight, ability to form disulfide interactions and sequence of their coding genes (Coleman and Larkins, 1998; Holding and Messing, 2013). The presence of internal tandem variable repeats with blocks of amino acids (mostly proline and glutamine) in all maize zein proteins was shown to be a common characteristic of zein proteins (Geraghty et al., 1981). These zeins have been classified into subfamilies α, β, γ, and δ- Among these four, the β, γ, and δ-zeins have a higher proportion of sulphur-rich amino acids. The δ-zeins are rich in Met (22%), whereas the γ-zeins are abundant in cysteine (Cys); β-zeins have a high percentage of Met and Cys (11%), while α-zeins lack Met and Cys (Wu et al., 2012). The α-zeins (19 and 22 kDa), containing a higher amount of proline and glutamine, account for 50% of total zeins, and are the reason for lower Met, Lys and Trp in maize kernels (Wu et al., 2009). The γ-zeins (50 kDa, 27 kDa and 16 kDa) contain 6.48%, 7.84% and 9.20% Met, respectively. The β- (15 kDa) and δ- (18 and 10 kDa) zeins are relatively Met rich and account for 11%, 27% and

22% of kernel Met, respectively (Table 1) (Swarup et al., 1995). The 18 kDa δ-zein have maximum Met among all zeins, because it is formed by the duplication of 10 kDa δ-zein through the allotetraploidization process (Swigonova et al., 2004; Wu et al., 2009). Along with this this δ-zein gene also contains one codon for Lys and two codons for Trp, which is not present in the mature 10 kDa δ-zein gene. Because of this characteristic, 18 kDa δ-zein has a higher nutritional value (Wu et al., 2009).

1.1.2 General regulation of zein genes

The most abundant α-zeins (19 and 22 kDa) are synthesized by four highly duplicated gene families distributed across six chromosomal locations on 4S consisting of more than 40 genes (Feng et al., 2009). Gene families Z1A, Z1B and Z1D encode the 19 kDa α-zeins, whereas Z1C sub-family encodes 22 kDa α-zeins (Song et al., 2001; Song and Messing, 2002; Feng et al., 2009). On the contrary, the β- (15 kDa), γ- (16, 27 and 50 kDa) and δ- (10 and 18 kDa) zeins are encoded by single-copy genes (Xu and Messing, 2008). The 15 kDa β-zeins are encoded by the *z2β15* gene. The 50, 27 and 16 kDa γ-zeins are encoded by *z2γ50*, *z2γ27* and *z2γ16*, respectively. The 10 kDa and 18 kDa δ-zeins are encoded by *z2δ10* and *z2δ18* genes on chromosomes (ch) 9 and 6, respectively (Xu and Messing, 2008; Liu et al., 2016). Due to their high levels of expression and complexity, zein synthesis serves as a model system to analyze the coordinated genetic regulation of several genes expressed at a specific developmental stage (Soave and Salamini, 1984).

1.1.3 Dzr1: a mutation that increases grain methionine concentration

Expression and accumulation of the 10 kDa δ-zein are regulated post-transcriptionally by the delta zein regulator (*dzr1*) which stabilizes the transcript of *dzs10*. This transcript has 22.5% codons for Met (Kirihara et al., 1988; Cruz-Alvarez et al., 1991; Schickler et al., 1993; Chaudhuri and Messing, 1995; Lai and Messing, 2002). It is also reported that the target sequence of the *dzr1* regulator is located in the untranslated regions (UTRs) of the mRNA of the *dzs10* gene (Lai and Messing, 2002). In maize two different regulation patterns of *dzr1* were reported. One regulatory pattern was derived from the genome and the other was derived from the environment *i.e.* dependent on its parental origin (Chaudhuri and Messing, 1995). In a study conducted by Olsen et al. (2003) when the B101 Met-rich line was crossed with other inbred lines or utilized as a male Met-donor parent, the *dzs10* allele was lost. This suggested that several other genetic factors, apart from *dzr1*, are involved in the expression of *dzs10*. Lai and Messing (2002) developed a Met-enriched transgenic maize line by replacing the cis-acting site (Mo17 allele) of *dzr1* regulation. The removal of the binding site of *dzr1* in *dzs10* uplifts the transcriptional control and resulted in high Met accumulation (Lai and Messing, 2002).

The 18 kDa δ-zein is regulated differently than that of the 10 kDa δ-zein. The UTRs of both the genes are different, to the extent that they regulate their mRNA accumulation through different RNA-protein interactions. The UTRs of 18 kDa δ-zein lack the prolamin box which is present in the 10 kDa δ-zein promoter sequence, which leads to different transcriptional levels in both the zeins (Wu et al., 2009).

TABLE 1 Methionine composition in zein protein of different classes (Wu et al., 2012).

Zeins	Molecular weight (kDa)	*Methionine (%)
α	22	0.00
	19 A	0.94
	19 B	0.46
	19 D	0.46
γ	50	1.08
	27	0.49
	16	1.84
β	15	11.25
δ	18	25.26
	10	22.48

*Methionine (%) = % methionine in total seed protein.

1.2 Methionine: its source, function and importance in the animal and human sector

1.2.1 Role of dietary methionine in animal systems

Methionine can be procured from both plant- and animal-based products. It is involved in a variety of metabolic functions and plays a vital part in cellular processes. The three most important functions are: 1) trans-methylation to form S-adenosylmethionine (SAM), a primary methyl donor that methylates compounds to form products like creatine, which is used to produce energy in the form of adenosine triphosphate (ATP) during vigorous exercise, and phosphatidylcholine, which is a structural component of cell membranes mostly present on the outer leaflets of membranes and also plays a role in cell signaling; 2) trans-sulphuration to produce Cys, which is then incorporated into glutathione (used to protect against oxidative damage under oxidative stress) or catabolized to taurine (Martinez et al., 2017); and 3) protein synthesis using a pool of Cys, Thr, and Ile derived from protein breakdown. Methionine also has a role in collagen formation, which is a structural component of connective tissues including skin and cartilages. It also aids the liver in metabolizing fat, preventing its storage, as well as protecting arteries from fat accumulation. It acts as a sulphur supplier to the body (Toohey, 2014). Ruan et al. (2017) have extensively reviewed the role of Met in animal immunity. Seymour (2016) have reviewed the role of Met in the diets of transition dairy cows. Methionine has been largely seen in the context of milk protein and fat synthesis as a co-limiting amino acid, however, now it is considered important for supporting the liver function and oxidative balance, besides immunity (Martinez et al., 2017). The dairy cow has increased metabolic and immunological requirements during the stage of calving. Methionine is associated with metabolic balance of proteins,

lipids and antioxidants (Pedernera et al., 2010). Cardoso et al. (2021) showed the positive impact of metabolizable protein and rumen-protected Met on the nutritional and immune status of lactating cows. Junior et al. (2021) demonstrated that a correct balance of Met to Lys promotes optimal utilization of amino acids in lactating cows.

Robinson and Bertolo (2016) have reviewed the role of Met in neonates. The authors conclude that Met is highly essential and the only indispensable sulphur amino acid. Methionine availability was being affected by a number of factors, primarily trans-methylation to other molecules. They highlighted the role of dietary methyl donors in ensuring a regular supply of Met. Garlick (2006) reviewed the toxicity of Met in the human body. Normal fluctuations in dietary Met are well tolerated, however, an increase of five times the normal requirement can potentially enhance homocysteine levels, which can increase susceptibility to cardiovascular disease. In infants also, a 2-5 times higher intake can result in impaired growth, however, no long-term consequences are observed. A recent study highlighted the importance of Met metabolism in human longevity (Mota-Martorell et al., 2021). Methionine supplementation enhanced mitochondrial pyruvate uptake and tricarboxylic acid (TCA) cycle activity. The authors implicate enhanced Met trans-sulfuration to be associated with longevity in humans. Disturbances in Met metabolism are also associated with a number of path physiologies. Singhal et al. (2018) implicated decrease in circulating levels of Met in the onset of multiple sclerosis. Siener et al. (2016) demonstrated that L-Met reduces the risk of struvite and calcium phosphate stone in healthy people.

On the other hand, Alachkar et al. (2022) have implicated the role of Met in neuroinflamation. Similarly, Tapia-Rojas et al. (2015) mentioned L-Met as the trigger for Alzheimer's disease. Pi et al. (2021) have found 5-methylcytosine as an associated factor in high dietary Met-induced Alzheimer's disease. Similarly, a number of studies indicate the importance of Met restriction as a therapeutic option in cancer biology (Wanders et al., 2020). Yu et al. (2018) have studied short-term Met deprivation as a strategy to reduce body fat, restore normal weight and glycemic control in mice. Navik et al. (2021) have described Met as a double-edged sword considering its varied role in human metabolism. Neubauer and Landecker (2021) have extensively reviewed the life cycle assessment of synthetic Met and have strongly argued for a reconsideration of the public health effects of anthropogenic augmentation of Met in food supply.

Hence, there's a need to consider the impact of dietary Met on public health in order to frame administrative policies around it. It is also essential to carefully assess the source of Met while measuring its potential effect on health. Plant proteins are taken up by animals, which later enter the food chain. Biological food processing and cooking results in changes to the protein structure and function. Oxidation of the dietary protein has been linked to a number of diseases (Estevez and Luna, 2017). It has also been linked to aging and age-related disorders (Stadtman, 2001). Dominguez et al. (2021) reviewed the issue of protein oxidation in muscle foods. The sulphur containing amino acids, Met and Cys are the most easily oxidizable amino acids. The irreversible oxidative modification of essential amino acids negatively impacts both their bioavailability as well as potential nutritional benefits. He et al. (2018) demonstrated that meat processing results in protein oxidation and aggregation due to changes in surface hydrophobicity and protein secondary structure. Similarly, Soladoye et al. (2015) have reviewed protein oxidation in processed meats. The authors report increase in freed radical generation and decrease of antioxidant activity in cooked meat, both of which contribute to protein oxidation. Methionine is known to be vulnerable to reactive oxygen species (Dorta et al., 2019; Dominguez et al., 2021). Sultana et al. (2009) have demonstrated that Met at position 35 in β-amyeloid protein is highly critical for associated neurotoxicity. This warrants a necessity to study the impact of source of Met and its association in the pathophysiology of Alzheimer's disease.

It is suggested that along with the bio-fortification, incorporation of antioxidant system in maize will be beneficial for animal feed (Xiong and Guo, 2020). The authors mention that plants contain many antioxidant compounds like tocopherols, polyphenols etc. Several studies suggest that the consumption of antioxidant rich plant food reduce the relative oxidation rate of proteins and lipids in the animal system (Faustman et al., 1998). Although, it is demonstrated that feeding antioxidant-rich plant extracts to chickens improved the oxidative stability of muscle lipids, but not proteins (Smet et al., 2008). Deb-Choudhury et al. (2020) demonstrated that myofibrillar proteins in animal tissue, like myosin, troponin and collagen are highly prone to oxidative protein modifications. Protein oxidation is correlated to the processing severity of food prepration (Estrada, 2019) and need to be taken into account, while accounting for the potential health effects of Met. Recently, He et al. (2022) demonstrated that Met oxidation is responsible for metastasis of pancreatic tumor cells. Given this information, the source of Met (in the form of plant or animal protein) also needs to be carefully considered while re-assessing the Met life cycle as suggested by Neubauer and Landecker (2021). Sato et al. (2021) have demonstrated the potential of amino acid-enriched plant-based therapeutic food for restoring amino acids levels in plasma of malnourished people. With the increasing concern for environmental and health-related issues, it is necessary to make a thorough scientific assessment of the current knowledge to guide administrative policies around human food components.

1.2.2 Ideal sulphur-containing amino acid: methionine or cysteine?

Methionine is an ideal sulphur-containing amino acid, and the reason for choosing Met is that it is an essential amino acid, whereas Cys is not. Another reason for choosing Met over Cys is that Met metabolism solely provides the total sulphur amino acid (TSAA) requirement of the body through the Met trans-sulphuration pathway, as Met acts as a precursor for Cys. Whereas, Cys is not able to act as a precursor for Met because the Met trans-sulphuration pathway is irreversible (Wheeler and Latshaw, 1981; Finkelstein et al., 1988; Finkelstein, 1990; Baker et al., 1996). A number of experiments on Cys as a Met sparing molecule (Wheeler and Latshaw, 1981; Finkelstein et al., 1988; Baker et al., 1996) have been undertaken in which Met was replaced by Cys in the trans-sulphuration pathway and found to be ineffective. The reason

behind this is raising Cys and reducing Met in the diet resulted in rising of dietary organic sulphur at the same concentration of total sulphur amino acids. These results suggested that the sparing of Met *via* Cys may alter the quantity of organic sulphur in TSAA (Chung and Baker, 1992; Fatufe, 2004).

1.3 Overview of methionine biosynthesis and metabolism

1.3.1 Sulphate reduction and assimilation

Amino acids are synthesized in bundle sheath cells during seed development in plants and transferred to the seeds, where they are stored in the form of proteins that are compatible with seed dormancy and germination. Sulphur is taken up by plants from the soil, reduced, and then integrated into Cys and Met in the leaf bundle sheath cells through sulfur assimilation (Figure 1). Methionine is generated from Cys as an intermediary step and other molecules including SAM, sulfolipids, and glutathione that also help to reduce the sulphur. Sulfate molecules are reduced to form Cys, and excess of Cys flow for the synthesis of Met. The Cys acts as a donor of thiol moiety to form Met, which indicates that the sulphur uptake and its reduction are the two key steps which are involved in biosynthesis of sulphur containing amino acids in plants (Wu et al., 2012).

Sulfate (SO_4^{2-}) is the stable form of sulphur present in the soil, and plants can takeup this sulfate *via* roots, and convert +6 oxidation state of sulfate to −2 oxidation state of organic sulphur with the help of three enzymes, ATP sulphurylase, adenosine phosphosulfate reductase (APR) and sulfite reductase. The rate of sulfate assimilation is relatively low, accounting for 5% of the rate of nitrate assimilation and 0.1-0.2% of photosynthetic carbon reduction. The activities involved in sulfate assimilation are so minute that it is difficult to elucidate the reaction mechanism involved. The enzymes involved in Cys biosynthesis have been found in the cytosol, plastids and mitochondrion of various plants which reflects the inability of these organelles to transport Cys across their membranes (Wu et al., 2012). SO_4^{2-} is first converted to adenosine 5'-adenylylsulfate (APS) through the enzyme ATP sulphurylase, APS then is converted to SO_3^{2-} by the enzyme APR, then to sulfide (S^{2-}) via sulfite reductase enzyme. Sulfide (S^{2-}) molecule reacts with the O-acetyl serine (OAS) and forming Cys through the enzyme OAS thiol-lyase (OAStL). O-acetyl serine is formed from serine and acetyl CoA through serine acetyltransferase (SAT) enzyme. Formed Cys molecules are used for the synthesis of Cys rich proteins (β- and γ-zeins) and Met, Met then incorporation into the Met-rich δ- zeins protein (Figure 1).

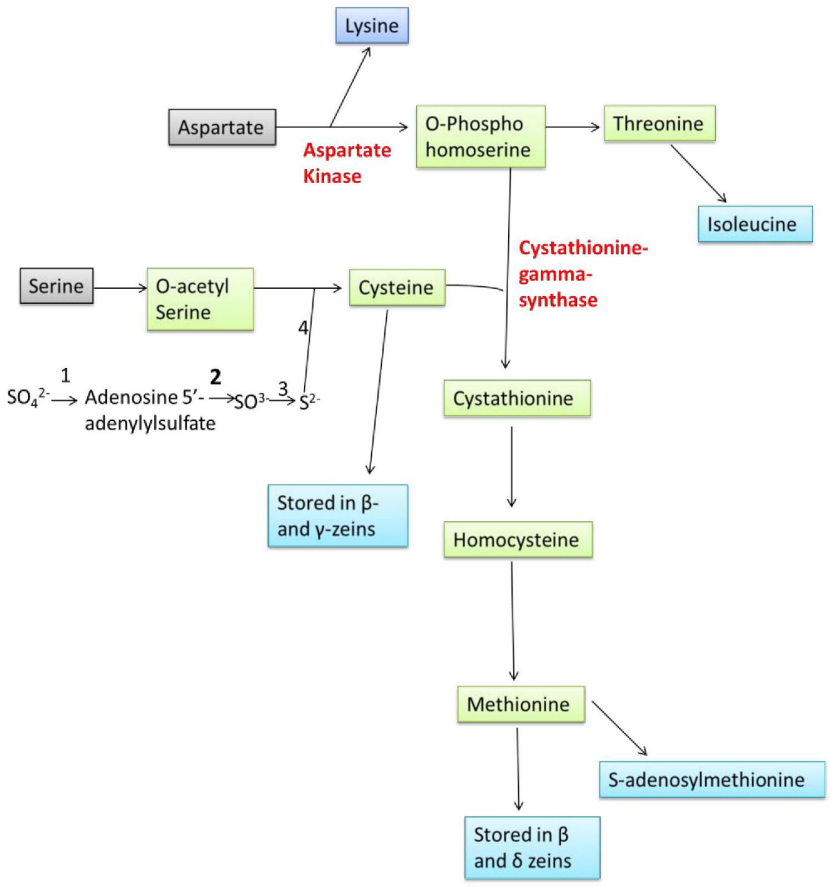


FIGURE 1
Overview of sulphur assimilation, methionine synthesis and its storage (1: ATP Sulfurylase, 2: APS Reductase, 3: Sulfite Reductase, 4: OAS thiol-lyase).

1.3.2 Biosynthesis of methionine

The aspartate family of amino acids consists of four members *viz.*, Met, Thr, Lys and Ile. These amino acids are formed from a common precursor aspartate (Figure 2). Aspartate kinase (AK) is the first enzyme that converts aspartate into β-aspartyl semialdehyde. The two amino acids, Thr and Lys accumulation inhibits the AK enzyme through feedback inhibition; thus, their excess quantity regulates Met synthesis. β-aspartyl semialdehyde has two fates, either it goes towards the Lys synthesis or it is converted to O-phosphohomoserine (OPH) *via* homoserine. The OPH either goes towards the synthesis of Thr by the enzyme threonine synthase (TS), or it goes towards the synthesis of Met. Hence, OPH acts as common substrate for both cystathionine-gamma-synthase (CGS) and TS. These two enzymes (TS and CGS) compete with each other to form their respective products (Zeh et al., 2001). Flux of carbon is higher towards the Thr synthesis when Met and SAM contents are higher. SAM content leads to the activation of threonine synthase enzyme. Methionine is formed by three step process from OPH. The Cys and OPH is converted to cystathionine by CGS, cystathionine is converted to homocysteine and then Met using cystathionine-β-lyase (CbL), and methionine synthase (MS), respectively (Hesse and Hoefgen, 2003).

1.3.3 Regeneration of methionine from SAM (SAM Cycle)

Methionine synthesis takes place in the bundle sheath cells of the leaves and accumulates there during the vegetative phase. It is converted to S-methyl-L-methionine (SMM) through the activity of methionine S-methyl transferase and then transported to the seeds in the form of SMM or other appropriate forms during reproductive phase (Bourgis et al., 1999; Bartlem et al., 2000; Kim et al., 2002). The SMM are converted back to the Met in the seeds *via* the enzyme homocysteine S-methyl transferase (HMT) (Ranocha et al., 2001). There are many isoforms of HMT but HMT3 is highly expressive in the seed. The isotope labeling of SMM and Met have shown that SMM produced in rosette leaves is translocated through the stem to the seeds (Cohen et al., 2017). The importance of SMM permeases in the phloem channel (Tan et al., 2010) and role of SMM in maize

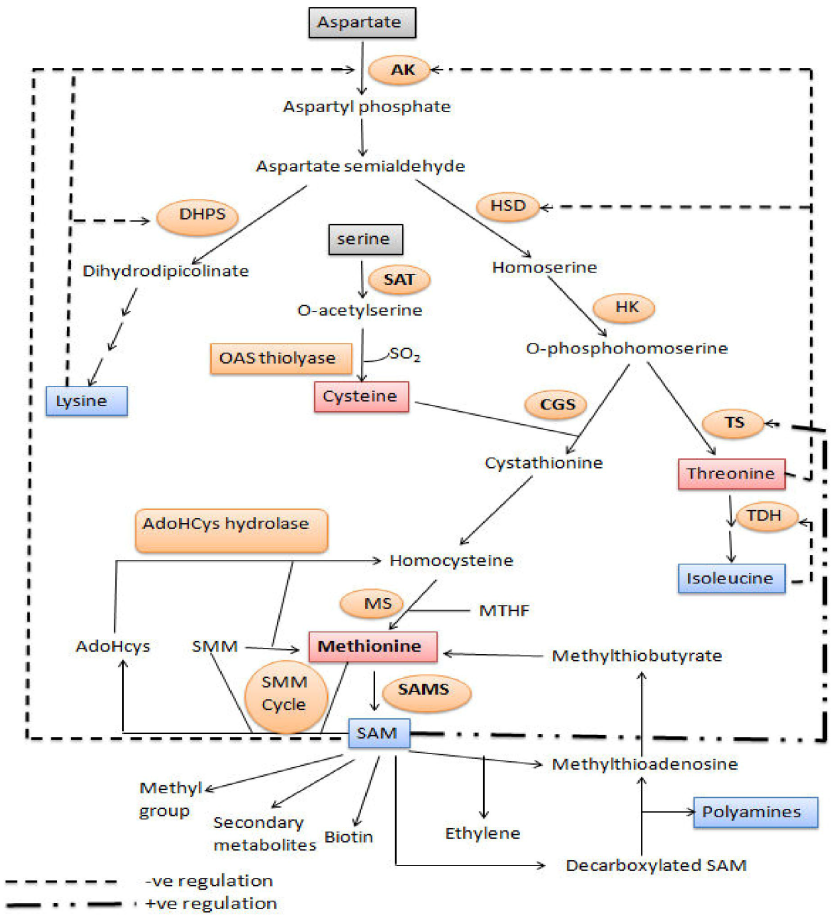


FIGURE 2
Methionine biosynthesis and metabolism: Diagram of metabolic pathway showing aspartate family pathway, metabolism of methionine and cysteine biosynthesis. Dashed arrows represent feedback inhibition of key enzymes. The bold dashed and dotted arrow represents the stimulation of threonine synthase activity through SAM. (Amir, 2008). DHPS, dihydrodipicolinate synthase; AK, aspartate kinase; HK, homoserine kinase; TS, threonine synthase; HSD, homoserine dehydrogenase; SAT, serine acetyl transferase; TDH, threonine dehydratase; OAS (thio) lyase, O-acetyl serine (thio) lyase; CBL, cystathionine-β-lyase; CGS, cystathionine-γ-synthase; SAM, S-adenosyl methionine; MS, methionine synthase; AdoHcys, adenosylhomocysteine; SAMS, S-adenosyl methionine synthase; SMM, S-methylmethionine; MTHF, methyltetrahydrofolate.

mmt mutants (having insertionally inactivated methionine S-methyl transferase enzyme) is not fully understood because *mmt* mutants and wild type *Arabidopsis thaliana* plants both have the same amount of Met (Kocsis et al., 2003). The transport form of Met at both early (Amir et al., 2019) and late-stage (Gallardo et al., 2007) of seed development is still under investigation in different crops.

It is reported that around 20% of the Met goes towards the synthesis of proteins, while the rest of it is converted to SAM *via* S-adenosylmethionine synthase. Decarboxylated SAM is formed through SAM decarboxylation. It supplies an aminopropyl group during polyamine production, following which Met is reproduced. The SAM is a biologically significant biomolecule because it plays a role in a variety of vital plant processes, including acting as a methyl group donor for DNA methylation and contributing to the discrimination of daughter and parent DNA during DNA repair. It also serves as a secondary metabolite and aids in the production of biotin and ethylene (Amir, 2008).

1.4 Genetic understanding of methionine content in maize

1.4.1 Quantitative Trait Loci associated with methionine in maize

In the future, improving Met content in maize kernels will rely on identifying genomic areas and gene networks that are consistently linked to Met production and accumulation. It is necessary to plan experiments on a diverse population with inherent natural variation in order to spot the important genetic areas.

Grain quality is an important feature that regulates the breeding programmes. The essential genes regulating the composition and structure of endosperm have been studied extensively. Multiple modifier genes contribute to superior grain quality under different genetic backgrounds (Ciceri et al., 2000; Huang et al., 2004). The marker-assisted breeding strategy has great potential to enhance the efficiency of the selection of desired traits (Babu et al., 2005). Quality Protein Maize (QPM) breeding programmes can be accelerated by using markers associated with essential amino acid, *opaque2 (o2)* and kernel hardness. Several studies were directed to identify QTLs associated with endosperm modification and protein quality including Met levels. Rojas et al. (2010) screened parental lines for differences in Met content; along with it, they identified four QTLs associated with Met content which together explained 57.3% variation among genotypes. Essential alleles that enhanced Met content were contributed by B73O2 inbred line. Among four reported QTLs, two QTLs were located near QTLs for Met, Lys and Trp-related traits. A QTL cluster on chromosome Ch7L is essentially associated with genes regulating Met and Trp synthesis. Along with it, a QTL cluster on Ch8S contains QTLs for three essential amino acids (Met, Trp, Lys). Both QTL clusters showed 73.9% phenotypic variance for the level of these three amino acids indicating the importance of clusters in the overall protein content of maize. Breeders are more interested in genes and QTLs that are directly involved in amino acid synthesis (metabolic pathways), such as four QTLs governing Met accumulation that were linked to genes regulating Met synthase (Lawrence et al., 2007). Similarly, the anthranilate synthase gene was located with the QTLs involved in phenotypic variation for Met, Trp and Lys accumulation (Lawrence et al., 2007).

1.4.2 Additional genetic components that regulate the methionine levels in maize

Kernel texture is a key feature in bio-fortified maize, along with essential amino acid synthesis in a balanced way. The hard kernel texture is negatively correlated with Lys content. It is established that endosperm modification in *o2* lines decreases the Lys accumulation in contrast to unmodified counterparts. This decrease in Lys content in modified kernels is justified by increased zein content (Robutti et al., 1974).

Additionally, the Met and Trp accumulation also show an inverse relationship. In *o2* genotypes with modified endosperm, Trp content was higher than Met as compared to normal maize inbred (Bantte and Prasanna, 2004; Scott et al., 2004). Amino acids Lys and Met are synthesized *via* a similar metabolic route, with aspartate (Asp) as a common precursor (Azevedo et al., 2006), whereas the chorismate pathway is involved in Trp production which is also crucial for secondary metabolite production (Radwanski and Last, 1995). Although numerous research has been conducted to better understand metabolic processes for amino acid synthesis, very little is known about the genetic regulation of these metabolic pathways in maize (Azevedo et al., 2006). Free amino acids (FAA) are another important nutrition quality trait and their QTLs are reported on Ch2L, Ch2S, Ch3S and Ch7L. Later 11 QTLs for free amino acid content were reported on chromosome 10 (Wang and Larkins, 2001). A gene for AK (*Ask2*) was found to be linked to a QTL on Ch2L affecting FAA content (Wang et al., 2007). In another study, among six reported QTLs for Met content, two QTLs on chromosome 2 were essentially associated with genes regulating Met content in maize. The rest were located as one QTL in Ch3, two QTLs in Ch4 and one in Ch8. The four QTLs associated with Met/ATT (aspartate-derived amino acid family related traits) were located on Ch1, 2 and 8, whereas QTLs for Met/Total trait were clustered on Ch1, 2, 3, 6 and 7 (Deng et al., 2017).

In multiple studies, several QTLs on different chromosomes are mapped for increasing amino acid content in the endosperm. The QTL cluster identified for Trp and Met was mapped close to the *o2* gene; however, although *o2* should not be segregated in tested population, it is unlikely to be associated with QTLs, which leaves AK and a 27 kDa γ-zein gene to be potential candidates found in this region (Robutti et al., 1974; Azevedo et al., 2006; Lawrence et al., 2007; Rojas et al., 2010).

1.4.3 QTLs associated with methionine in orthologous cereal crop

Rice, like maize, is the focus of multiple studies aimed at increasing Met content and its linked genetic regions. Research published in 2002 found 80 QTLs for 19 amino acid content in milled rice, with QTLs for Met content accounting for over 7.4% of the variation (Wang et al., 2008). The 48 and 64 QTLs and 12 QTL clusters for amino acid accumulation were elucidated by Zhong et al. (2011) including qAa1 on Ch1, qAa7 on Ch7 and qAA9 on Ch9. Among these clusters, qAa1 and qAa9 negatively and qAa7

positively regulated the amino acid accumulation. Thus, the QTLs specifically affecting Met content were located on ch1, 9 and 7. In another study, a total of 17 QTLs for amino acid content were reported on six chromosomes *viz.* 1, 3, 6, 7, 8 and 10. Five of the 17 QTLs, qAAC6.1, qAAC6.2, qAAC7.1, qAAC7.2, and qAAC8.2, were shown to be linked to important amino acid genes. Certain QTLs (qAAC6.1 and qAAC7.1) from the reported study had multiple effects on more than one amino acid. Biosynthesis of almost 11 amino acids (Alanine (Ala), Arginine (Arg), Aspartate (Asp), Glutamic acid (Glx), Glycine (Gly), Isoleucine (Ile), Leucine (Leu), Met, Phenylalanine (Phe), Tyrosine (Tyr), and Valine (Val)) was contributed by QTL qAAC6.1 (Jang et al., 2020).

According to the compilation of various studies in rice, it can be concluded that QTLs majorly affecting amino acid accumulation are located on eight chromosomes *viz.* 1, 2, 3, 4, 6, 7, 8, 10. A total of 5 QTL clusters on chromosomes 1, 2, 7 and 8 which majorly affect amino acid accumulation, specifically the region of 1.5-5.19 Mb on Ch1 that has 4 QTLs for amino acid content (Wang et al., 2007; Zhong et al., 2011; Yoo, 2017; Jang et al., 2020). The genetic variations are controlled by QTLs, but none of these have yet been cloned (Demidov et al., 2003). Therefore, the identification of relevant QTLs and their effect on the Met content may lead to developing new maize inbreds with higher Met storage proteins which can be utilized to target specific genomic regions that have been conserved over time for amino acid accumulation and protein quality attributes. The studies on QTLs associated with Met in maize and rice crops are listed in Table 2.

The present review's compiled information on QTL can be employed in meta-QTL analysis to investigate the congruency of the identified regions related to certain traits of interest. Regions associated in rice could be used for identifying hotspot genomic regions in maize. The information from the compiled meta-QTL studies will be used for the development of trait-specific KASP (kompetitive allele-specific PCR) markers which will be an important stepping stone to accelerating the maize nutritional quality program globally.

1.5 Dissection and understanding of proteome rebalance to enhance methionine level

The proportion of essential amino acids (Met, Lys) in maize kernels is determined by the type of maize protein. In previous investigations, high δ-zein lines were found to be sufficient to substitute synthetic Met in a complete feed for poultry and human consumption (Messing and Fisher, 1991). However, for the development of maize lines having a higher content of both Lys and Met, a dissection of the mechanism underlying proteome rebalancing is required. Accumulation of Met-rich δ-zein was associated with the post-transcriptional regulation of its mRNA (Schickler et al., 1993; Lai and Messing, 2002), whereas high-Lys maize lines (*o2* mutant) depend on the compensatory increase of Lys-rich non-zeins. According to the mechanism, an increase in Lys level tends to be accompanied by a decrease in Met content (Mertz et al., 1964; Phillips and Mcclure, 1985; Wu et al., 2012). The decreased Met level in Lys-rich *o2* mutant is related to *o2* mediated reduction of β- and δ-zeins. The proteomic analysis also reports that only 8% of proteins in maize kernel possess Met above 4%, whereas, 57% of proteins have Lys above 4% (Wu et al., 2012). As a result, the mechanisms for accumulating these two amino acids differ throughout seed development, and both must be targeted simultaneously in order to develop nutritionally superior maize lines. Another problem with targeting Met-rich lines is its linkage with a reduced level of Cys-rich γ-zein (27 kDa) (Newell et al., 2014). These studies suggest that increased content of Met during seed storage needs higher reduced sulphur flow towards Met synthesis from Cys, which resulted in the reduced translation of Cys-rich γ-zein (27 kDa) mRNA. However, the development of QPM with high Lys, Trp and vitreous kernel texture depends on the increased expression of 27 kDa γ-zein. Overall, proteome reframing is needed to increase the essential amino acids Lys, Try and Met without compromising kernel texture and agronomic performance.

2 Strategies for development of methionine-rich maize lines

2.1 Selection of methionine-rich inbred and natural mutant lines

In conventional breeding, the Met concentration in different inbred lines varies greatly. Screening of maize genotypes for high Met percentage followed by suitable breeding program in order to fix this trait along with the better agronomic performance is required to enhance the overall quality. Three cycles of screening and recurrent selection of maize lines can be used to sort the germplasm into different groups based on desired traits. As described in the previous study (Sethi et al., 2021), the temporal expression of δ-zein after pollination can be evaluated to determine the differential pattern of protein accumulation during seed development. It indicates the developmental stage at which maize seeds have maximum Met content that can be used as selection criteria for screening of Met-rich lines. The utilization of grains at that particular developmental stage having maximum Met may help to make the availability of more amount of Met to animals and humans (Sethi et al., 2021). According to earlier reports, the B101 is characterized as a Met-rich line. The parental lines of "Vivek Hybrid 9" (CM145 and CM212) were changed to QPM version through transfer of the *o2* gene with the help of MAS and phenotypic screening for endosperm modifications. It's QPM version was named "Vivek QPM 9" and released in India during 2008. There was significant improvement of Met by 3.4% in the "Vivek QPM 9" hybrid compared to its normal version "Vivek Hybrid 9" (Gupta et al., 2009; Gupta et al., 2013).

Natural mutants with high Met concentration, such as floury-2 (*fl2*), were identified about 55 years ago (Nelson et al., 1965). The *fl2*

TABLE 2 List of detail of QTLs linked with methionine, essential amino acids, and total amino acid content in maize and rice.

Traits	QTL name	Chr.	Marker/interval (Mbp)	LOD value	R^2	References
Maize						
Methionine	*qMet5*	5	bnlg1046	3.76	8.9	Rojas et al., 2010
	qMet7	7	umc2142	8.91	20.7	
	qMet8	8	umc1304	5.07	18.3	
Methionine	*qMet2*	2	PZE-102080745	4.09	8.84	Deng et al., 2017
	qMet2.1	2	SYN1715	3.78	7.88	
	qMet3	3	PZE-103099938	4.38	8.93	
	qMet4	4	SYN2317	3.53	8.49	
	qMet4.1	4	PZE-104138961	4.15	8.70	
	qMet8	8	PZE-108087230	3.76	8.95	
Methionine/ATT	*qMet/ATT1*	1	ZM013506-0433	4.38	10.33	
	qMet/ATT2	2	PZE-102080745	4.97	10.65	
	qMet/ATT8	8	PZE-108087230	3.32	7.69	
Methionine/Total amino acids	*qMet/Total1*	1	ZM013506-0433	3.02	6.52	
	qMet/Total2	2	PZE-102080745	5.01	10.86	
	qMet/Total3	3	SYN37729	4.63	9.09	
	qMet/Total3.1	3	PZE-103075963	4.09	7.96	
	qMet/Total3.2	3	PUT-163a-101384606-6	3.78	7.59	
	qMet/Total6	6	PZE-106006541	3.57	8.49	
	qMet/Total7	7	SYNGENTA3646	3.42	7.29	
Free amino acids	*qFAA*	2L	bmc1633- bmc1329	—–	11	Wang and Larkins, 2001
	qFAA	2S	bmc1537-bmc2248	—–	10	
	qFAA	3S	bmc1904-bmc2136- bmc1452	—–	15	
	qFAA	7L	bmc2328b- phi045	—–	10	
Rice						
Ala/Arg/Asx/Glx/Gly/Ile/Leu /Met/Phe/Tyr/Val	*qAAC6.1*	6	3.59	16.9	10.81	Jang et al., 2020
Methionine	*qAAC6.2*	6	4.21	9.3	5.23	
Total amino acid	*qAa1*	1	RM493-RM562 (12.3-14.6)	12.3	24.2	Zhong et al., 2011
	qAa9	9	RM328-RM107 (19.7-20.1)	8.1	13.2	
Eaa/Total/Asp/Thr/Glu/Gly/Ala/Cys/Tyr/Pro	*qAA.1*	1	RM472-RM104 (37.9-40.2)	—–	—–	Wang et al., 2007
Pro/Gly/Met/Arg	*qAA.7*	7	RM125-RM214 (5.5-12.8)	—–	—–	

encodes unusual α-zeins leading to aberrant protein body formation and an opaque phenotype (Lending and Larkins, 1992). Thereafter, a maize germ line BSSS53 having higher Met content was discovered, as evidenced by higher expression of Met-rich storage protein-producing genes (δ-zein) (Phillips and Mcclure, 1985). Mutations in the genes for *aspartokinase1* (*ask1*) and *aspartokinase2* (*ask2*), encoding enzymes free from feedback inhibition by Lys, have been found in maize (Diedrick et al., 1990). The mutation in these loci resulted in enhanced production of Thr, Lys, Met and isoleucine (Dodson et al., 1990). It has been reported that a maize line Oh54502 with *ask2* mutant showed a higher accumulation of free amino acids in the endosperm (Wang et al., 2007).

2.2 Genetic engineering approaches to enhance the methionine content

Since there are a number of control points in and around the metabolic network of Met, conventional breeding may not be able to increase Met levels without compromising yield and quality parameters. As a result, genetic engineering may be a viable option for developing high Met maize lines

2.2.1 Modulating Sulphur-flux toward methionine synthesis

Recent studies suggested that enhancing the accumulation of total sulphur protein and 10 kDa δ-zein without any negative effect on other zeins can be attained by either enhanced S-assimilation in the leaf (Xiang et al., 2018) or decreasing sulfate reduction capacity in the maize kernel (Planta et al., 2017), although sulphur assimilation in kernel cannot be ruled out (Tabe and Droux, 2001). The key enzyme is APR which is not expressed in the kernel. Enhanced expression of two genes *APR* and *SAT* in bundle sheath cells increases the flux through the pathway, resulting in more Met and δ-zeins. However, the accumulation of toxic intermediates in the plants led to stunted growth (Martin et al., 2005), which was the major negative impact on the plant. If this hurdle could be solved, then S-assimilation might potentially enhance the resource for Met accumulation in maize seeds (Wu et al., 2012).

Another enzyme CGS catalyzed the synthesis of Met from Cys. If the activity of this CGS enzyme in maize lines can be increased, more Cys can be synthesized, diverting surplus Cys to Met production and, as a result, higher levels of Met-rich storage proteins in maize kernels can be reached. Amir and Tabe (2006) reported that highly expressive gene of *CGS* from *Arabidopsisthaliana* was transferred into transgenic tobacco (*Nicotiana tabacum*) and alfalfa (*Medicago sativa*) which leads to the accumulation of free Met content in leaves but does not show any effect on the storage protein-bound Met. But, opposite results were attained in transgenic tobacco seeds which express a truncated version of the *Arabidopsis cystathionine-γ-synthase* (*AtCGS*) gene. The feedback-insensitive form of CGS enzyme produces less soluble Met but more bound Met (Matityahu et al., 2013). Both soluble and bound Met was produced in higher amount in *Arabidopsis thaliana* (Cohen et al., 2014) and soybean (Cohen and Amir, 2017) seeds with the same *AtCGS* construct with legume promoter. The feedback insensitive CGS mutants accumulated Met and performed better under oxidative stress (Cohen et al., 2014). The role of carbon and nitrogenous substrates for Met content has recently been observed, where *aspartokinase* for carbon flow and *CGS* for nitrogen flow were modified with respect to their feedback sensitivity. The binary vector under control of suitable promoters has been transformed into tobacco seeds and leaves resulting in a 2 and 170-fold increase of protein-bound Met, respectively (Hacham et al., 2008; Matityahu et al., 2013). The HMT pulling of accumulated SMM in leaves of *Arabidopsis thaliana* plants with suppressed *CGS* through RNA interference led to 33% increase in seed Met (Cohen et al., 2017). The over-expression of truncated *CGS* in vegetative tissues of rice increased the activity but did not improve the Met content, which once again highlighted the incompetence of rice leaves to throw excess Met to the seed (Whitcomb et al., 2018).

In maize CGS, APR and SAT enzymes are the major key regulating enzymes for sulphur flux towards Met synthesis, so up-regulation of these enzymes will result in the enhanced sulphur flow, which finally affects the Met accumulation (Figure 3).

2.2.2 Enhancing the sulphur assimilation

The level of Met and zein fractions can be modulated by developing different mutant lines with altered enzymatic activities of sulphur assimilatory or Met biosynthesis pathways. To circumvent many control points, the bacterial *APR* homologs may be excellent candidates to develop Met rich transgenic lines. Under limiting sulphur amino acids (SAA) availability, the B101 exemplified the utmost natural threshold of Met accumulation in maize grain. This threshold apparently can be overcome with increased sulphur-reduction and assimilation during photosynthesis. The role of non-seed tissues in producing and transporting Met to the seed has been studied through genetic engineering. The transgenic maize line developed by using *AtSAT1* gene from *Arabidopsis thaliana* under the control of leaf bundle sheath cell-specific promoter *rbcS1* (Rubisco small subunit 1) exhibited 12-fold higher activity of SAT enzyme and led to higher sulphur assimilation (Xiang et al., 2018). Thus, 4-fold higher amount of soluble Met and Cys in maize lines without any negative impact on maize plants has been observed. Higher Met resulted due to a higher level of 10 kDa δ-zein storage protein, representing 1.40-fold higher protein-incorporated Met in maize kernel (Xiang et al., 2018). Similarly, constitutive over-expression of the *Escherichia coli SAT* gene (*EcSAT*) in rice resulted in a 40% and 4.8-fold increase in soluble and protein-bound Met, respectively (Nguyen et al., 2012). A 17% increase in the bound-Met and 74% increase in Cys of soybean seeds has been observed when *O-acetyl serine (thiol) lyase* gene (*OAS-tL*) was overexpressed (Kim et al., 2012). This finding again confirms the threshold limit of Met accumulation under the presence of pooled Cys. Recently it is reported that the highest Met containing line was developed by stacking the two transgene (*pRbcs : AtSAT1-pRbcs : EcPAPR*), which resulted in 2.24 fold increase of Met in maize (Xiang et al., 2022). Hence, engineering the *SAT* gene of maize for enhanced expression results in higher sulphur assimilation, which positively influences the Met accumulation in the seed (Figure 3).

2.2.3 Modulating feedback inhibition

Genetic engineering gives the opportunity to make carbon or sulphur flow-biosynthetic enzymes resistant/insensitive to feedback inhibition, allowing it to produce more enzymes even when feedback inhibitors are present. The chimeric genes producing the Met-rich *2S* albumin of Brazil nuts (*Bertholletia excelsa*) and bacterial feedback insensitive *aspartate kinase* gene have been inserted into narbon bean (*Vicia narbonensis*) (Saalbach et al., 1995; Demidov et al., 2003) and rapeseed (*Brassica napus*) (Altenbach et al., 1992) to improve the amino acid balance of

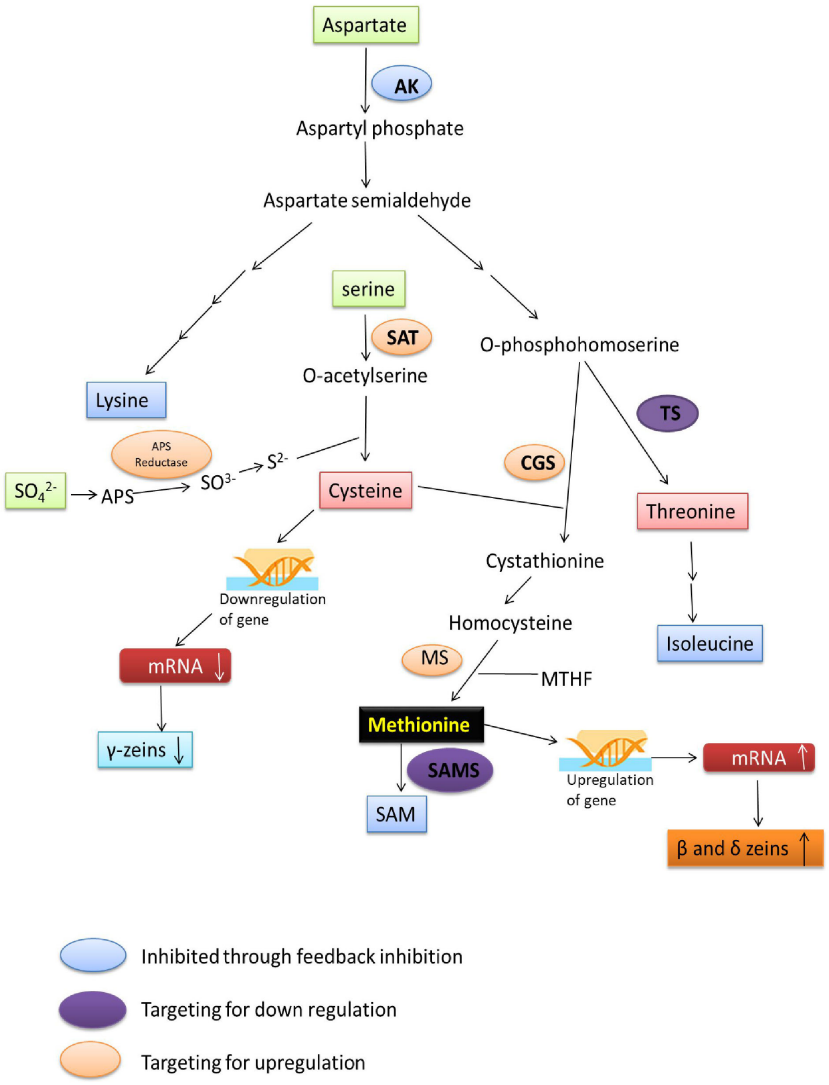


FIGURE 3
Flow diagram representing target key enzymes for genetic engineering to enhance Met expression. AK, aspartate kinase; SAT, serine acetyl transferase; APS Reductase, adenosine 5-adenylylsulfate Reductase; APS, adenosine 5-adenylylsulfate; CGS, cystathionine-γ-synthase; TS, threonine synthase; MS, methionine synthase; SAMS, S-adenosyl methionine synthase.

seeds. In maize identifying the different isoforms of the AK insensitive to feedback inhibition and enhanced expression of genes of sulphur assimilation, Met biosynthesis and AK pathway, such as *SAT*, *APR*, *AK*, *CGS* and *MS* could be the potential way to enhance Met accumulation (Figure 3).

2.2.4 Overexpression of seed storage proteins

Another approach to enhance amino acid levels is by replacing a storage protein that is lacking in an amino acid with a protein that is enriched with particular amino acid. The sink capacity could be enhanced through the expression of genes encoding Met-rich protein content in the seed. Enhanced expression of Met-rich maize proteins probably led to lower expression of γ- and β-zeins genes (Wu and Messing, 2010). Silencing of genes of γ- and β-zeins proteins showed that these were the main sink of Cys amino acid which act as a donor of sulphur for Met synthesis (Wu et al., 2012).

The higher Met level could be possible at the salvage of other sulphur-containing metabolites and/or enzymes. The reduced synthesis of Cys-containing proteins leads to increased sulphur flux to other amino acids such as Met, resulting in higher δ-zein proteins in the seed (Wu et al., 2012; Wu and Messing, 2014). In the condition of suppression of endogenous sulphur-poor proteins and expression of γ-zeins, available sulphur plays a significant role in the accumulation of Met-rich storage protein (Kim et al., 2014). The expression of sulphur-rich proteins in growing maize seeds was accompanied by a decrease in endogenous sulphur-rich proteins. This suggests that the sulphur shortage triggered the sulphur-protein reallocation (Lai and Messing, 2002; Hagan et al., 2003; Chiaiese et al., 2004). In maize, the seeds having less γ-zeins proein, accumulate 10-56% more Met (Taylor et al., 2008; Yin et al., 2011; Newell et al., 2014). Surprisingly, RNAi-induced gene silencing for genes of β- and γ-zeins proteins has no impact on Met content.

Since, Met receives its sulphur mostly from Cys, it appears that the ratio of these two amino acids influences the protein synthesis, and their protein-bound content is regulated by translational altitude (Wu et al., 2012). Recent study showed that the expression of Met-rich maize gene encoding for β-zein protein into soybean under the control of legumin B4 promoter or the CaMV 35S promoter gives the high Met containing transgenic soybean seeds, in which the Met content is significantly increased by up to approximately 15% (Guo et al., 2020). Enhancing expression of Met rich maize delta zein, including 10 kDa δ-zein which is associated with the *dzs10* gene (Zm00001eb382030) located in chr9:47915348-47917942bp and 18 kDa δ-zein associated with the *dzs18* gene (Zm00001eb281380) located in chr6:125301025-125303844bp (MaizeGDB, https://www.maizegdb.org/data_center/map) are key targets to enhance the Met percentage in maize.

In conclusion, targeting the maize protein with a lower Met percentage such as α, β and γ-zeins for its inhibition, or enhancement of protein rich in Met (10 and 18 kDa δ-zeins) will be the potential way for enhancing the protein-bound Met (Figure 3).

2.2.5 Overexpression of methionine biosynthesis-related enzymes

Studies revealed that transgenic tobacco seeds expressing bacterial insensitive *aspartokinase* under the seed-specific promoter at the final stage of seed development produce 6.5% more bound Met (Karchi et al., 1993) whereas, overexpression in vegetative tissues led to no significant change in seed Met (Hacham et al., 2008). The latter study suggests that the higher level of Met in leaves does not push the excess into developing seeds.

The enzyme methionine gamma lyase (MGL) is present abundantly in the cytoplasm of all the plant's organs except seeds which converts Met into methanethiol, ammonia and α-ketobutyrate. It has been reported that this enzyme had a higher k_m value for the Met, indicating that the enzyme performed its function when there is higher Met accumulation in the cell (Huang et al., 2014). The higher accumulation of Met and SMM content has been observed in the mutants of *Arabidopsis thaliana* lacking MGL. Over-expression of the *AtCGS* in potato tubers, along with the RNAi : MGL construct, resulted in a 2.2-fold increase in Met (soluble form) accumulation in comparison to the wild-type (Kumar and Jander, 2017). It has also been observed that the MGL enzyme performs its activity under certain conditions such as sulphur starvation, and leads to the degradation of Met (Ufaz and Galili, 2008). The reported studies suggest that the genes of aspartokinase, CGS and MGL enzymes could be targeted to increase the Met percentage *via* genetic engineering techniques.

2.2.6 Gene silencing to enhance methionine

Using a transgenic approach, the levels of protein-bound Met could be enhanced by approximately 30–97% in comparison to their non-transgenic form (Altenbach et al., 1992; Molvig et al., 1997; Lai and Messing, 2002; Lee et al., 2003; Newell et al., 2014). The transgenic *Arabidopsis thaliana* (*mto2*, Methionine over-accumulator) and potato plants with deficient TS enzyme activity produced higher free Met content in leaves (Bartlem et al., 2000; Zeh et al., 2001). The silencing of miRNA reduced the expression of *TS* gene which may lead to the availability of O-homoserine for Met synthesis resulting in enhancement of the Met content of maize endosperm. This has been explained in a way that the enhanced Thr content inhibits the AK enzyme thus inhibiting the synthesis of Met, Lys and Ile. So diverting O-homoserine away from Thr biosynthesis may also prevent the inhibitory effect of Thr. Additionally, accumulation of SAM stimulates the activity of TS enzyme, which inhibits the Met synthesis (Hacham et al., 2002). The Inhibition of the activity of the SAM synthase enzyme may result in higher Met content in maize lines. It has been reported that, the *Arabidopsis thaliana* mutant *mto3* with lower expression of *SAM synthase* gene, resulted in up to 200 folds more Met content. These mutant lines did not show any visible growth difference from the wild type, except for a slight delay in germination (Shen et al., 2002).

Transgenic maize PE5 (contains the expression cassettes PepC-EcPAPR) with *Escherichia coli* gene *3′-phosphoadenosine-5′-phosphosulfate reductase* (*EcPAPR*) under the control of leaf-specific promoter, down-regulates the endogenous *APR* and increases the expression of gene coding for Cys-rich non-zein proteins. These transgenic maize lines produced 57.6% more kernel-bound Met than the best Met-rich maize inbred line B101 (Planta et al., 2017). Transgenic PE5 (maternal) even backcrossed twice to B101 prior to being crossed with α, β, γ, γ/α or γ/β- zein RNAi lines (paternal) produced a higher amount of Met. The loss of β- and γ-zeins, as compared to decreased α-zeins, increased the δ-zeins and produced vitreous kernels only in presence of PE5 i.e. PEF:γ (Wu et al., 2012; Planta et al., 2017). However, some decreases in 27 kDa γ-zeins were attributed to gene segregation (Segal et al., 2003) or gene silencing (Huang et al., 2004). Reduction in both β- and γ-zeins mobilizes even more protein sulphur to the 10 kDa δ-zein than the loss of either β- or γ-zeins individually. δ-zein (10 kDa) seems to be the most responsive to enhanced assimilative sulphate reduction followed by the 15 kDa β-, 16 kDa γ-, and 27 kDa γ-zein, respectively. This order also follows the number of sulphur amino acids in these zeins. Therefore, it seems that the higher the SAA residues of the zeins, the more responsive it would be to increased sulphur supply (Planta et al., 2017).

In maize, targeted inhibition of TS and SAM synthase enzymes will be beneficial to enhance the Met accumulation. Silencing the expression of sulphur-deficient maize proteins may result in the accumulation of sulphur-rich proteins to balance the overall protein percentage and that can be a better approach for developing Met-rich maize germplasm (Figure 3).

2.2.7 Modifying gene codons or promoters

By modifying the genes encoding proteins with relatively higher Met in such a way that their encoding regions contain more codons for Met, or engineering the promoter region of those genes to enhance their expression in turn leads to higher accumulation of δ-zein protein in maize endosperm during seed development. This strategy mainly encountered problems of protein instability and relatively small improvement (Krishnan, 2005; Beauregard and Hefford, 2006; Wenefrida et al., 2009). A gene coding for 11 kDa

δ-zein protein from maize germ line W23a1 was isolated and introduced into the soya bean plants. These transgenic lines showed 1.5 to 1.7% higher Met content than the non-transgenic lines (Kim and Krishnan, 2004). Till now no modified maize storage proteins are available. According to a study, the expression of the 18 kDa δ-zein protein encoding gene is minimal in most maize plants. Alteration in UTRs of the 18 kDa δ-zein encoding gene may lead to the higher accumulation of Met-rich proteins due to increased gene expression. Recently Planta et al. (2017) used this strategy with *E. coli* PAPS (3'-phosphoadenosine-5'-phosphosulfate) reductase (PPAR) and maize δ-zein and improved the Met content for 58% as compared to control without apparent proteome rebalancing. Thus, targeting the δ-zein protein or δ-zein encoding gene is important for Met regulation. The existence of alternative routes for Met synthesis through PPAR provides flexibility in metabolic engineering in maize.

2.2.8 Expression of synthetic genes

This strategy can be used to generate transgenic higher Met-maize lines expressing codon-optimized synthetic genes having kernel-specific promoters. It will help to transform Met, to change an imbalanced composition into a balanced one, and for that non-conserved regions of proteins must be replaced with a sequence high in Met. In sweet potato (*Ipomoea batatas*) and soybean, seed-specific expression of synthetic genes (*MB-16* for soybean) resulted in seed protein with 13-16 percent (higher) Met residues (Slater et al., 2008; Zhang et al., 2014). De Clercq et al. (1990) constructed the Met-rich *AT2S1* gene by replacing the poorly conserved region with Met-rich sequences of *2S* albumin gene 1 of *Arabidopsis thaliana* and the insertion of that construct in *Arabidopsis thaliana*, Brazil nut and tobacco resulted into Met-enriched 2S albumins in all of them. The new chimeric gene *zeolin* was designed to encode Met-rich proteins through fusion of maize γ-zein with bean storage protein phaseolin (Mainieri et al., 2004). This strategy has scope to form stable chimeric proteins with the required content of essential amino acids. The identification of non-conserved/non-functional regions from the maize genome and replacement of them with Met-rich sequence could be an alternative option to generate Met-rich transgenic maize lines.

2.3 Targeting identified candidate genes associated with methionine *via* genetic engineering/gene editing

Currently, gene editing is a potential method for altering a particular region of interest. In this work, a comprehensive list of genes (source MaizeGDB) associated with Met production has been prepared (Table 3) which can be used to enhance its level via various genetic engineering approaches. The 25 loci (S No. 36-60 in Table 3) were found to be associated with Met content using GWAS experiments across two environments in the previous study (Deng et al., 2017) which can be used for molecular screening of maize germplasm for higher Met content. Some of the genes for overexpression *via* genetic engineering are also listed (Table 4). It is well known that for gene editing Met-rich lines there is a need to target the negative regulators of Met production. Therefore, the list has been prepared for prospective candidate genes which can be knockout to develop transgene-free Met-rich genome-edited (GEd) maize lines (Table 4) *via* CRISPR-Cas9-based gene-editing technique.

TABLE 3 Details of candidate genes associated with methionine metabolism, which codes for enzymes, transcription factors, signaling and transporter proteins.

S. No.	Candidate gene	Metabolite	Chr	Annotation	Position (bp)	Reference
1.	GRMZM2G024686	Met	1	Aspartate kinase, conserved site	298313666-298414802	Wen et al., 2014
2.	GRMZM2G406746	Met	7	Pentatricopeptide repeat	9769078-9872377	
3.	GRMZM2G110145	Met	10	Cellulose synthase	77271453-77305285	
4.	GRMZM2G128319	Met	10	Protein kinase-like	113791164-113887667	
5.	GRMZM2G121275	Met	2	Major intrinsic protein	19780391-20248092	
6.	GRMZM2G121223	Met	2	Lipid-binding START	19780391-20248092	
7.	GRMZM2G702954	Met	2	Unknown	19780391-20248092	
8.	GRMZM2G101290	Met	2	Glucose/ribitol dehydrogenase	19780391-20248092	

(Continued)

TABLE 3 Continued

S. No.	Candidate gene	Metabolite	Chr	Annotation	Position (bp)	Reference
9.	GRMZM2G101181	Met	2	Protein of unknown functions DUF246, plant	19780391-20248092	
10.	GRMZM2G078648	Met	2	Zinc finger, FYVE/PHD-type	19780391-20248092	
11.	GRMZM2G172936	Met	2	Pathogenesis-related transcriptional factor and ERF, DNA-binding	19780391-20248092	
12.	GRMZM2G411639	Met	2	Class I peptide chain release factor	19780391-20248092	
13.	GRMZM2G028969	Met	2	Pathogenesis-related transcriptional factor and ERF, DNA-binding	19780391-20248092	
14.	GRMZM2G325513	Met	2	Antifreeze protein, type I	19780391-20248092	
15.	AC207188.3_FG002	Met	2	Unknown	19780391-20248092	
16.	GRMZM2G105137	Met	2	SANT, DNA-binding	19780391-20248092	
17.	GRMZM2G139458	Met	2	Unknown	19780391-20248092	
18.	GRMZM2G139463	Met	2	Peptidase T2, asparaginase 2	19780391-20248092	
19.	GRMZM2G024389	Met	1	Unknown	298313666-298414802	
20.	GRMZM2G024267	Met	1	Unknown	298313666-298414802	
21.	GRMZM2G024374	Met	1	Kinesin, motor region	298313666-298414802	
22.	GRMZM2G023242	Met	1	Nascent polypeptide-associated complex NAC	298313666-298414802	
23.	GRMZM2G022269	Met	1	Protein synthesis factor, GTP-binding	298313666-298414802	
24.	GRMZM2G022248	Met	1	Glutaredoxin-related protein	298313666-298414802	
25.	GRMZM2G420772	Met	7	Plant disease resistance response protein	1262063-1309782	
26.	GRMZM2G420743	Met	7	Plant disease resistance response protein	1262063-1309782	
27.	GRMZM2G420758	Met	7	Unknown	1262063-1309782	
28.	GRMZM2G420733	Met	7	Plant disease resistance response protein	1262063-1309782	
29.	GRMZM2G120652	Met	7	Vitamin B6 biosynthesis protein	1262063-1309782	
30.	GRMZM2G120575	Met	7	Protein of unknown function DUF1665	1262063-1309782	
31.	GRMZM2G120574	Met	7	Tyrosine protein kinase	1262063-1309782	
32.	GRMZM2G120572	Met	7	Unknown	1262063-1309782	
33.	GRMZM2G555108	Met	7	WD40 repeat	1262063-1309782	
34.	GRMZM2G120563	Met	7	C2 calcium/lipid-binding region, CaLB	1262063-1309782	
35.	GRMZM2G109627	Met	8	No apical meristem (NAM) protein	1262063-1309782	
36.	GRMZM2G095631 (chr4.S_46137424)	Met	4	Unknown	46137424	Deng et al., 2017

(Continued)

TABLE 3 Continued

S. No.	Candidate gene	Metabolite	Chr	Annotation	Position (bp)	Reference
37.	GRMZM2G160541 (chr4.S_143379704)	Met	4	Phenylalanine ammonia-lyase	143379704	
38.	GRMZM2G139412 (chr5.S_151841984)	Met	5	Shikimate kinase	151841984	
39.	GRMZM2G362298 (chr6.S_74144428)	Met	6	Acyl-transferase family protein	74144428	
40.	GRMZM2G351239 (chr1.S_190429440)	Met/AAT	1	Unknown	190429440	
41.	GRMZM2G159145 (chr3.S_13532873)	Met/AAT	3	Thioredoxin H-type	13532873	
42.	GRMZM2G160541 (chr4.S_143379704)	Met/AAT	4	Phenylalanine ammonia-lyase	143379704	
43.	GRMZM2G065451 (chr4.S_237774932)	Met/AAT	4	Transcription factor family protein	237774932	
44.	GRMZM2G425728 (chr7.S_9707633)	Met/AAT	7	Early light-induced protein	9707633	
45.	GRMZM2G015534 (chr7.S_10695002)	Met/AAT	7	Opaque endosperm2	10695002	
46.	GRMZM2G344911 (chr7.S_88939234)	Met/AAT	7	2,3-dihydro-2,3-dihydroxybenzoate dehydrogenase	88939234	
47.	GRMZM2G133806 (chr8.S_164189020)	Met/AAT	8	Unknown	164189020	
48.	GRMZM2G091819 (chr10.S_16572309)	Met/AAT	10	Disulfide oxidoreducatse	16572309	
49.	GRMZM2G351239 (chr1.S_190429440)	Met/Total	1	Unknown	190429440	
50.	GRMZM2G013283 (chr2.S_50503804)	Met/Total	2	Glutamate-ammonia ligase	50503804	
51.	GRMZM2G158316 (chr4.S_185735097)	Met/Total	4	Unknown	185735097	
52.	GRMZM2G425728 (chr7.S_9707633)	Met/Total	7	Early light-induced protein	9707633	
53.	GRMZM2G045834 (chr7.S_28399595)	Met/Total	7	Unknown	28399595	
54.	GRMZM2G133806 (chr8.S_164189020)	Met/Total	8	Unknown	164189020	
55.	GRMZM2G008226 (chr1.S_218566308)	Met	1	trehalose biosynthetic process	218566308	
56.	GRMZM2G166713 (chr4.S_165967277)	Met	4	methionine-tRNA ligase	165967277	
57.	GRMZM2G178106 (chr9.S_153267892)	Met	9	Beta-galactosidase	153267892	
58.	GRMZM2G175463 (chr9.S_153269528)	Met	9	Unknown	153269528	
59.	GRMZM2G176141 (chr.S_180848700)	Met/AAT	2	Transcription factor	180848700	
60.	GRMZM2G138727 (chr7.S_120199552)	Met/Total	7	Glutelin-2 Precursor (Zein-gamma) (27 kDa zein)	120199552	

*ATT=Aspartate-derived amino acid family related traits, Total= Total amino acids.

TABLE 4 Details of proposed genes to be targeted for the genetic engineering and genome editing for methionine.

S. No.	Candidate gene	Chr	Position (bp)	Annotation	Strategy	Reference
1.	Zm00001eb064530 (*ask1*)	1	305262144-305265827	Aspartate kinase	Knockout	Bartlem et al., 2000; Zeh et al., 2001; Avraham et al., 2005; Kumar and Jander, 2017; Hou et al., 2022; MaizeGDB
2.	Zm00001eb094670 (*ask2*)	2	161226305-161232000	Aspartate kinase		
3.	Zm00001eb392050 (*CGS1*)	1	65940342-659460321	Cystathionine-γ-synthase	Over-expression	
4.	Zm00001eb018300	1	65198840-65205925	Cystathionine-γ-synthase		
5.	Zm00001eb156060	3	212214968-212222772	Threonine synthase	Knockout	
6.	Zm00001eb294790 (*thr1*)	6	174066198-174071347	Threonine synthase		
7.	Zm00001eb284240 (*thr2*)	6	143594078-143599133	Threonine synthase		
8.	Zm00001eb022690 (*thr3*)	1	87258448-87262005	Threonine synthase		
9.	Zm00001eb088230	2	107810019-107813351	Threonine synthase		
10.	Zm00001eb361640 (*sat1*)	8	160277043-160280293	Serine acetyltransferase	Over-expression	
11.	Zm00001eb008110 (*sat2*)	1	24415858-24419122	Serine acetyltransferase		
12.	Zm00001eb293180 (*sat3*)	6	163558296-163561657	Serine acetyltransferase		
13.	Zm00001eb002740 (*sat4*)	1	7638621-7653175	Serine acetyltransferase		
14.	Zm00001eb322440 (*aprl1*)	7	157694136-157699277	PAPS Reductase	Down-regulation	
15.	Zm00001eb105800 (*aprl2*)	2	209131699-209136961	PAPS Reductase		
16.	Zm00001eb194120 (*aprl3*)	4	183478812-183483064	PAPS Reductase		
17.	Zm00001eb255320 (*aprl4*)	5	214156431-214161113	PAPS Reductase		
18.	Zm00001eb042300 (*aprl5*)	1	221287901-221308523	PAPS Reductase		
19.	Zm00001eb410330 (*aprl6*)	10	25080671-25087625	PAPS Reductase		
20.	Zm00001eb137590 (*aprl7*)	3	133490011-133494737	PAPS Reductase		
21.	Zm00001eb061530 (*aprl8*)	1	295632186-295637314	PAPS Reductase		
22.	Zm00001eb212580 (*aprl9*)	5	3625345-3630121	PAPS Reductase		
23.	Zm00001eb177530 (*akh1*)	4	68460416-68515471	Homoserine dehydrogenase	Over-expression	
24.	Zm00001eb097580 (*akh2*)	2	177654848-177677022	Homoserine dehydrogenase		

(Continued)

TABLE 4 Continued

S. No.	Candidate gene	Chr	Position (bp)	Annotation	Strategy	Reference
25.	Zm00001eb382030 (*dzs10*)	9	47915348-47917942	10kDa δ-zein	Over-expression	
26.	Zm00001eb281380 (*dzs18*)	6	125301025-125303844	18kDa δ-zein		

2.3.1 Combinatorial strategy

As methionine regulation is a complex process the strategies have to be used in combination to enhance the Met content. This combinatorial strategy works on improved sink and source relationship. In sink capacity, Met enriched albumin protein was added whereas the source was improved through making carbon family enzyme insensitive. Similar combinatorial strategy with insensitive SAT and sunflower albumin in lupin increased the Met content to 2-folds but the dosage effect of SAT was negligible (Tabe et al., 2010). Recently, in soybean, although the introduction of β-zein gene improved total Met content in the seeds, this level was negligible compared to native soybean storage proteins, implying that the inadequate soluble Met is the limiting factor. From these findings it is clear that the Met regulation is a multi-step process, and to up-regulate, a combined strategy needs to be designed for simultaneous increase of the source and sink of the Met metabolism. Various approaches to enhance the Met level in maize seed are represented in Figure 4.

3 Conclusion and future directions

The enhancement of Met in the seed for balancing the amino acid content is a tricky process. A number of metabolic enzymes, transporters, genes and transcription factors are involved in regulating the biosynthesis and accumulation of Met. Apart from

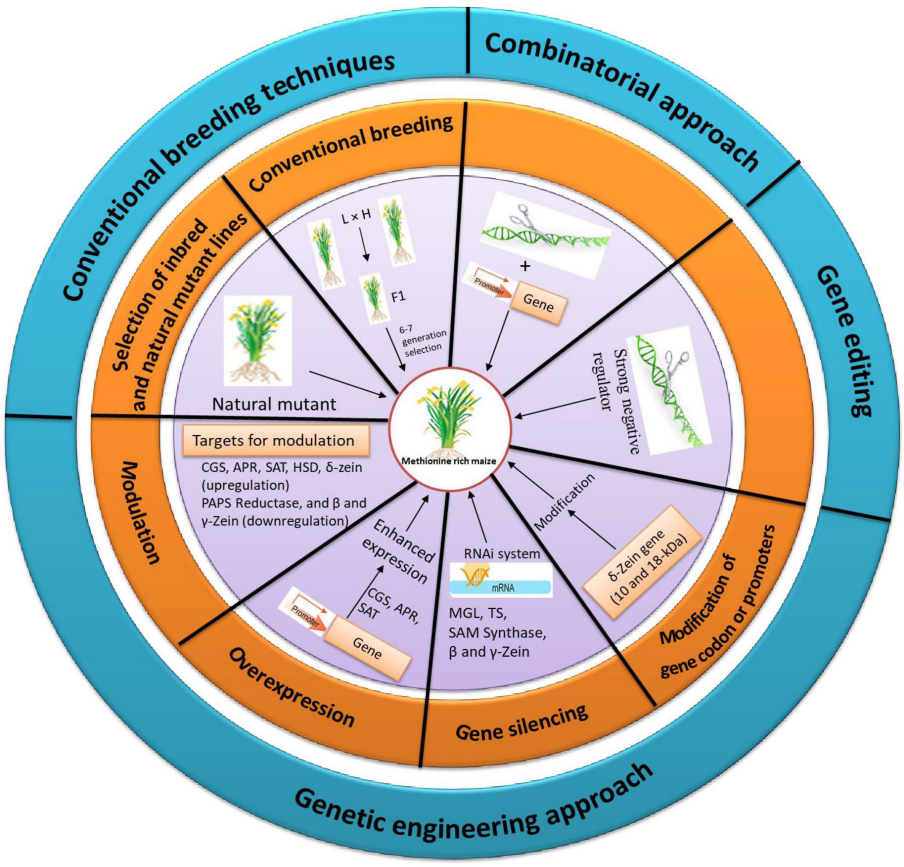


FIGURE 4
Different approaches and targets for the enhanced accumulation of methionine in maize.

it, temporal separation for Met transporting molecules and spatial separation of sulphur assimilation makes the control of whole biosynthetic pathway a difficult task. The tissue-specific expression of Cys in bundle sheath cells and glutathione in mesophyll cells compartmentalizes the sulphur at the next higher level of cellular hierarchy. Hence, a combination of multiple approaches without compromising the yield is required to be adopted. The genomic regions associated with the trait identified in orthologous cereal crops could be used for identifying hotspot genomic regions in maize for the trait of interest. This review provides trait-related information which can be used further in metaQTL analysis to study the congruency of the identified regions associated with specific traits of interest. The strategies explained in this review will give potential insight into target enzymes and candidate genes, which can be used to enhance Met content *via* genetic engineering or genome-editing approaches. Although these strategies seem to be promising, their effectiveness depends upon the additional studies to define the effect of these manipulations on traits such as seed morphology, seed starch, amino acids, oil content and germination rate. Thus, genetic manipulation had proven successful in increasing the content of Met in seeds. Nevertheless, supplemental information is required to confirm the *in vivo* stability of genetically engineered Met-rich proteins for optimal fortification of seeds. Once nutritionally improved genetically modified (GM)/genetically edited (GEd) genotypes possessing high Met content in their seeds are being developed, there will be an urgent need to analyze the genotype with the best response for superior agronomic traits to perform well under field conditions.

Author contributions

Conceptualization, DC and VD; writing-original draft preparation, VD; writing-review and editing, BB, VS, MS, AS, CK, MG, SR, RK, and DC; supervision, DC. All authors contributed to the article and approved the submitted version.

Funding

The authors declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

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

Publisher's note

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

References

Alachkar, A., Agrawal, S., Baboldashtian, M., Nuseir, K., Salazar, J., and Agrawal, A. (2022). L-methionine enhances neuroinflammation and impairs neurogenesis: Implication for Alzheimer's disease. *J. Neuroimmunol.* 366, 577843. doi: 10.1016/j.jneuroim.2022.577843

Altenbach, S. B., Kuo, C. C., Staraci, L. C., Pearson, K. W., Wainwright, C., Georgescu, A., et al. (1992). Accumulation of a Brazil nut albumin in seeds of transgenic canola results in enhanced levels of seed protein methionine. *Plant Mol. Biol.* 18, 235–245. doi: 10.1007/BF00034952

Amir, R. (2008). Towards improving methionine content in plants for enhanced nutritional quality. *Func. Plant Sci. Biotechnol.* 2, 36–46.

Amir, R., Cohen, H., and Hacham, Y. (2019). Revisiting the attempts to fortify methionine content in plant seeds. *J. Exp. Bot.* 70, 4105–4114. doi: 10.1093/jxb/erz134

Amir, R., and Tabe, L. (2006). "Molecular approaches to improving plant methionine content," in *Plant genetic engineering, vol 8: metabolic engineering and molecular farming II*. Eds. K. J. Pawan, K. Jaiwal and J. Rana (Houston, Texas: Studium Press, LLC), 1–26.

Avraham, T., Badani, H., Galili, S., and Amir, R. (2005). Enhanced levels of methionine and cysteine in transgenic alfalfa (Medicago sativa L.) plants over-expressing the Arabidopsis cystathionine γ-synthase gene. *Plant Biotechnol. J.* 3 (1), 71–79. doi: 10.1111/j.1467-7652.2004.00102.x

Azevedo, R. A., Lancien, M., and Lea, P. J. (2006). The aspartic acid metabolic pathway, an exciting and essential pathway in plants. *Amino Acids* 30, 143–162. doi: 10.1007/s00726-005-0245-2

Babu, R., Nair, S. K., Kumar, A., Venkatesh, S., Sekhar, J. C., Singh, N. N., et al. (2005). Two-generation marker-aided backcrossing for rapid conversion of normal maize lines to quality protein maize (QPM). *Theor. Appl. Genet.* 111, 888–897. doi: 10.1007/s00122-005-0011-6

Baker, D. H., Fernandez, S. R., Webel, D. M., and Parsons, C. M. (1996). Sulfur amino acid requirement and cystine replacement value of broiler chicks during the period three to six weeks posthatching. *Poult. Sci.* 75, 737–742. doi: 10.3382/ps.0750737

Bantte, K., and Prasanna, B. M. (2004). Endosperm protein quality and kernel modification in the quality protein maize inbred lines. *J. Plant Biochem. Biotechnol.* 13, 57–60. doi: 10.1007/BF03263192

Bartlem, D., Lambein, I., Okamoto, T., Itaya, A., Uda, Y., Kijima, F., et al. (2000). Mutation in the threonine synthase gene results in an over–accumulation of soluble methionine in arabidopsis. *Plant Physiol.* 123, 101–110. doi: 10.1104/pp.123.1.101

Beauregard, M., and Hefford, M. A. (2006). Enhancement of essential amino acid contents in crops by genetic engineering and protein design. *Plant Biotechnol. J.* 5, 561–574. doi: 10.1111/j.1467-7652.2006.00204.x

Bourgis, F., Roje, S., Nuccio, M. L., Fisher, D. B., Tarczynski, M. C., Li, C., et al. (1999). *S*-methylmethionine plays a major role in phloem sulfur transport and is synthesized by a novel type of methyltransferase. *Plant Cell.* 11, 1485–1497. doi: 10.1105/tpc.11.8.1485

Cardoso, F. F., Donkin, S. S., Pereira, M. N., Pereira, R. A. N., Peconick, A. P., Santos, J. P., et al. (2021). Effect of protein level and methionine supplementation on dairy cows during the transition period. *J. Dairy Sci.* 104, 5467–5478. doi: 10.3168/jds.2020-19181

Chaudhuri, S., and Messing, J. (1995). RFLP mapping of the maize dzr1 locus, which regulates methionine–rich 10 kDa zein accumulation. *Mol. Gen. Genet.* 246, 707–715. doi: 10.1007/BF00290716

Chiaiese, P., Ohkama-Ohtsu, N., Molvig, R., Godfree, H., Dove Hocart, C., Fujiwara, T., et al. (2004). Sulphur and nitrogen nutrition influence the response of chickpea seeds to an added, transgenic sink for organic sulphur. *J. Exp. Bot.* 55, 1889–1901. doi: 10.1093/jxb/erh198

Chung, T. K., and Baker, D. H. (1992). Apparent and true amino acid digestibility of a crystalline amino acid mixture and of casein: comparison of values obtained with ileal-cannulated pigs and cecectomized cockerels. *J. Anim. Sci.* 70, 3781–3790. doi: 10.2527/1992.70123781x

Ciceri, P., Castelli, S., Lauria, M., Lazzari, B., Genga, A., Bernard, L., et al. (2000). Specific combinations of zein genes and genetic backgrounds influence the transcription of the heavy-chain zein genes in maize *opaque-2* endosperms. *Plant Physiol.* 124, 451–460. doi: 10.1104/pp.124.1.451

Cohen, H., and Amir, R. (2017). Dose-dependent effects of higher methionine levels on the transcriptome and metabolome of transgenic Arabidopsis seeds. *Plant Cell Rep.* 36, 719–730. doi: 10.1007/s00299-016-2003-1

Cohen, H., Hacham, Y., Matityahu, I., and Amir, R. (2017). "Elucidating the effects of higher expression level of cystathionine γ–synthase on methionine contents in transgenic arabidopsis, soybean and tobacco seeds," in *Sulphur metabolism in higher plants - fundamental, environmental and agricultural aspects. Proceedings of the International Plant Sulphur Workshop*. Eds. L. De Kok, M. Hawkesford, S. Haneklaus and E. Schnug (Cham: Springer), 39–48. doi: 10.1007/978-3-319-56526-2_4

Cohen, H., Israeli, H., Matityahu, I., and Amir, R. (2014). Seed–specific expression of a feedback-insensitive form of cystathionine-γ-synthase in Arabidopsis stimulates metabolic and transcriptomic responses associated with desiccation stress. *Plant Physiol.* 166, 1575–1592. doi: 10.1104/pp.114.246058

Coleman, C. E., and Larkins, B. A. (1998). *The prolamins of maize in seed proteins*. Eds. P. R. Shewry. and R. Casey (Dordrecht: Kluwer Academic Publishers), 109–139.

CROPS (2018). *Maize vision 2022 a knowledge report, Indian Maize Summit* (New Delhi: Agricultural Times, FICCI), 1–34. Available at: https://agritimes.co.in/market-reports/maize-vision-2022-India/.

Cruz-Alvarez, M., Kirihara, J. A., and Messing, J. (1991). Post-transcriptional regulation of methionine content in maize kernels. *Mol. Gen. Genet.* (New Delhi) 225, 331–339. doi: 10.1007/BF00269866

Darrigues, A., Buffard, C., Lamkey, K. R., and Scott, M. P. (2005). Variability and genetic effects for tryptophan and methionine in commercial maize germplasm. *Maydica* 50, 147–156.

Deb-Choudhury, S., Haines, S., Harland, D., Clerens, S., Van Koten, C., Lee, E., et al. (2020). Multi–parameter evaluation of the effect of processing conditions on meat protein modification. *Heliyon* 6, e04185. doi: 10.1016/j.heliyon.2020.e04185

De Clercq, A., Vandewiele, M., Van Damme, J., Guerche, P., Van Montagu, M., Vandekerckhove, J., et al. (1990). Stable accumulation of modified 2S albumin seed storage proteins with higher methionine contents in transgenic plants. *Plant Physiol.* 94, 970–979. doi: 10.1104/pp.94.3.970

Demidov, D., Horstmann, C., Meixner, M., Pickardt, T., Saalbach, I., Galili, G., et al. (2003). Additive effects of the feed–back insensitive bacterial aspartate kinase and the Brazil nut 2S albumin on the methionine content of transgenic narbon bean (*Vicia narbonensis* L.). *Mol. Breed.* 11, 187–201. doi: 10.1023/A:1022814506153

Deng, M., Li, D., Luo, J., Xiao, Y., Liu, H., Pan, Q., et al. (2017). The genetic architecture of amino acids dissection by association and linkage analysis in maize. *Plant Biotechnol. J.* 15, 1250–1263. doi: 10.1111/pbi.12712

Diedrick, T. J., Frisch, D. A., and Gengenbach, B. G. (1990). Tissue culture isolation of a second mutant locus for increased threonine accumulation in maize. *Theor. Appl. Genet.* 79, 209–215. doi: 10.1007/BF00225953

Dodson, M. V., Mathison, B. A., and Mathison, B. D. (1990). Effects of medium and substratum on ovine satellite cell attachment, proliferation and differentiation in *vitro*. *Cell Differ. Dev.* 29, 59–66. doi: 10.1016/0922-3371(90)90024-q

Dominguez, R., Pateiro, M., Munekata, P. E. S., Zhang, W., Garcia-Oliveira, P., Carpena, M., et al. (2021). Protein oxidation in muscle foods: A comprehensive review. *Antioxidants* 11, 60. doi: 10.3390/antiox11010060

Dorta, E., Avila, F., Fuentes-Lemus, E., Fuentealba, D., and Lopez-Alarcon, C. (2019). Oxidation of myofibrillar proteins induced by peroxyl radicals: Role of oxidizable amino acids. *Food Res. Int.* 126, 108580. doi: 10.1016/j.foodres.2019.108580

Estevez, M., and Luna, C. (2017). Dietary protein oxidation: A silent threat to human health? *Crit. Rev. Food Sci. Nutr.* 57, 3781–3793. doi: 10.1080/10408398.2016.1165182

Estrada, P. D. (2019). *Protein oxidation as an integrated aspect of plant protein-based food* (WU, Wageningen University: Wageningen University). doi: 10.18174/499098

Fatufe, A. (2004). *Dose response relationships between intake and efficiency of utilisation of individual amino acids in chicken* (Gottingen, Germany: Cuvillier Verlag Publishing), 164.

Faustman, C., Chan, W. K., Schaefer, D. M., and Havens, A. (1998). Beef color update: the role for vitamin E. *J. Anim. Sci.* 76, 1019. doi: 10.2527/1998.7641019x

Feng, L., Zhu, J., Wang, G., Tang, Y. P., Chen, H. J., and Jin, W. B. (2009). Expressional profiling study revealed unique expressional patterns and dramatic expressional divergence of maize α-zein super gene family. *Plant Mol. Biol.* 69, 649–659. doi: 10.1007/s11103-008-9444-z

Finkelstein, J. D. (1990). Methionine metabolism in mammals. *J. Nutr. Biochem.* 1, 228–237. doi: 10.1016/0955–2863(90)90070–2

Finkelstein, J. D., Martin, J. J., and Harris, B. J. (1988). Methionine metabolism in mammals: The methionine-sparing effect of cysteine. *J. Biol. Chem.* 263, 11750–11754. doi: 10.1016/S0021-9258(18)37847-5

Food and Agriculture Organization (2022). *Statistical yearbook* (United Nations: World food and agriculture).

Gallardo, K., Firnhaber, C., Zuber, H., Hericher, D., Belghazi, M., Henry, C., et al. (2007). A combined proteome and transcriptome analysis of developing *Medicago truncatula* seeds: Evidence for metabolic specialization of maternal and filial tissues. *Mol. Cell Proteom.* 6, 2165–2179. doi: 10.1074/mcp.M700171-MCP200

Garlick, P. J. (2006). Toxicity of methionine in humans. *J. Nutr.* 136, 1722S–1725S. doi: 10.1093/jn/136.6.1722S

Geraghty, D., Peifer, M. A., Rubenstein, I., and Messing, J. (1981). The primary structure of a plant storage protein: zein. *Nucleic Acids Res.* 9, 5163–5174. doi: 10.1093/nar/9.19.5163

Gibbon, B. C., and Larkins, B. A. (2005). Molecular genetic approaches to developing quality protein maize. *Trends Genet.* 21, 227–233. doi: 10.1016/j.tig.2005.02.009

Guo, C., Liu, X., Chen, L., Cai, Y., Yao, W., Yuan, S., et al. (2020). Elevated methionine content in soybean seed by overexpressing maize β-zein protein. *Oil Crop Sci.* 5, 11–16. doi: 10.1016/j.ocsci.2020.03.004

Gupta, H. S., Agrawal, P. K., Mahajan, V., Bisht, G. S., Kumar, A., Verma, P., et al. (2009). Quality protein maize for nutritional security: Rapid development of short duration hybrids through molecular marker assisted breeding. *Curr. Sci.* 96, 230–237.

Gupta, H. S., Raman, B., Agrawal, P. K., Mahajan, V., Hossain, F., and Thirunavukkarasu, N. (2013). Accelerated development of quality protein maize hybrid through marker–assisted introgression of *opaque-2* allele. *Plant Breed.* 132, 77–82. doi: 10.1111/pbr.12009

Hacham, Y., Avraham, T., and Amir, R. (2002). The N-terminal region of Arabidopsis cystathionine γ-synthase plays an important regulatory role in methionine metabolism. *Plant Physiol.* 128, 454–462. doi: 10.1104/pp.010819

Hacham, Y., Matityahu, I., Schuster, G., and Amir, R. (2008). Overexpression of mutated forms of aspartate kinase and cystathionine γ-synthase in tobacco leaves resulted in the high accumulation of methionine and threonine. *Plant J.* 54, 260–271. doi: 10.1111/j.1365-313X.2008.03415.x

Hagan, N. D., Upadhyaya, N., Tabe, L. M., and Higgins, T. J. V. (2003). The redistribution of protein sulfur in transgenic rice expressing a gene for a foreign, sulfur-rich protein. *Plant J.* 34, 1–11. doi: 10.1046/j.1365-313X.2003.01699.x

He, D., Feng, H., Sundberg, B., Yang, J., Powers, J., Christian, A. H., et al. (2022). Methionine oxidation activates pyruvate kinase M2 to promote pancreatic cancer metastasis. *Mol. Cell.* 82, 3045–3060.e11. doi: 10.1016/j.molcel.2022.06.005

He, J., Zhou, G., Bai, Y., Wang, C., Zhu, S., Xu, X., et al. (2018). The effect of meat processing methods on changes in disulfide bonding and alteration of protein structures: Impact on protein digestion products. *RSC Adv.* 8, 17595–17605. doi: 10.1039/C8RA02310G

Hesse, H., and Hoefgen, R. (2003). Molecular aspects of methionine biosynthesis. *Trends Plant Sci.* 8, 259–262. doi: 10.1016/S1360-1385(03)00107-9

Holding, D., and Messing, J. (2013). "Evolution, structure, and function of prolamin storage proteins," in *seed genomic*. Ed. P. W. Becraft (Oxford, UK: Wiley-Blackwell), 138–158. doi: 10.1002/9781118525524.ch8

Hou, S., Men, Y., Zhang, Y., Zhao, K., Ma, G., Li, H., et al. (2022). Role of miRNAs in regulation of SA-mediated upregulation of genes involved in folate and methionine metabolism in foxtail millet. *Front. Plant Sci.* 13, 1023764. doi: 10.3389/fpls.2022.1023764

Huang, S., Adams, W. R., Zhou, Q., Malloy, K. P., Voyles, D. A., Anthony, J., et al. (2004). Improving nutritional quality of maize proteins by expressing sense and antisense zein genes. *J. Agric. Food Chem.* 52, 1958–1964. doi: 10.1021/jf0342223

Huang, T., Joshi, V., and Jander, G. (2014). The catabolic enzyme methionine gamma-lyase limits methionine accumulation in potato tubers. *Plant Biotechnol. J.* 12, 883–893. doi: 10.1111/pbi.12191

Jang, Y. H., Park, J. R., and Kim, K. M. (2020). Antimicrobial activity of chrysoeriol 7 and chochlioquinone 9, white–backed plant hopper-resistant compounds, against rice pathogenic strains. *Biol. (Basel)* 9, 382. doi: 10.3390/biology9110382

Junior, V. C., Lopes, F., Schwab, C. G., Toledo, M. Z., and Collao-Saenz, E. A. (2021). Effects of rumen–protected methionine supplementation on the performance of high production dairy cows in the tropics. *PloS One* 16, e0243953. doi: 10.1371/journal.pone.0243953

Karchi, H., Shaul, O., and Galili, G. (1993). Seed–specific expression of a bacterial desensitized aspartate kinase increases the production of seed threonine and methionine in transgenic tobacco. *Plant J.* 3, 721–727. doi: 10.1111/j.1365-313X.1993.00721.x

Kim, W. S., Chronis, D., Juergens, M., Schroeder, A. C., Hyun, S. W., and Jez, J. M. (2012). Transgenic soybean plants overexpressing O-acetylserine sulfhydrylase accumulate enhanced levels of cysteine and Bowman-Birk protease inhibitor in seeds. *Planta* 235, 13–23. doi: 10.1007/s00425-011-1487-8

Kim, W. S., Jez, J. M., and Krishnan, H. B. (2014). Effects of proteome rebalancing and sulfur nutrition on the accumulation of methionine rich δ-zein in transgenic soybeans. *Front. Plant Sci.* 5. doi: 10.3389/fpls.2014.00633

Kim, W. S., and Krishnan, H. B. (2004). Expression of an 11 kDa methionine-rich delta-zein in transgenic soybean results in the formation of two types of novel protein bodies in transitional cells situated between the vascular tissue and storage parenchyma cells. *Plant Biotechnol. J.* 2, 199–210. doi: 10.1111/j.1467-7652.2004.00063.x

Kim, J., Lee, M., Chalam, R., Martin, M. N., Leustek, T., and Boerjan, W. (2002). Constitutive overexpression of cystathionine γ-synthase in Arabidopsis leads to accumulation of soluble methionine and *S*-methylmethionine. *Plant Physiol.* 128, 95–107. doi: 10.1104/pp.101801

Kirihara, J. A., Petri, J. B., and Messing, J. (1988). Isolation and sequence of a gene encoding a methionine-rich 10-kDa zein protein from maize. *Gene* 71, 359–370. doi: 10.1016/0378-1119(88)90053-4

Kocsis, M. G., Ranocha, P., Gage, D. A., Simon, E. S., Rhodes, D., Peel, G. J., et al. (2003). Insertional inactivation of the methionine *S*-methyltransferase gene eliminates

the *S*-methylmethionine cycle and increases the methylation ratio. *Plant Physiol.* 131, 1808–1815. doi: 10.1104/pp.102.018846

Krishnan, H. B. (2005). Engineering soybean for enhanced sulfur amino acid content. *Crop Sci.* 45, 454–461. doi: 10.2135/cropsci2005.0454

Kumar, P., and Jander, G. (2017). Concurrent overexpression of *Arabidopsis thaliana* cystathionine γ-synthase and silencing of endogenous methionine γ-lyase enhance tuber methionine content in *Solanum tuberosum*. *J. Agric. Food Chem.* 65, 2737–2742. doi: 10.1021/acs.jafc.7b00272

Lai, J., and Messing, J. (2002). Increasing maize seed methionine by mRNA stability. *Plant J.* 30, 395–402. doi: 10.1046/j.1365-313X.2001.01285.x

Lawrence, C. J., Schaeffer, M. L., Seigfried, T. E., Campbell, D. A., and Harper, L. C. (2007). Maize GDB's new data types, resources and activities. *Nucleic Acids Res.* 35, D895–D900. doi: 10.1093/nar/gkl1048

Lee, T. T. T., Wang, M. M. C., Hou, R. C. W., Chen, L. J., Su, R. C., Wang, C. S., et al. (2003). Enhanced methionine and cysteine levels in transgenic rice seeds by the accumulation of sesame 2S albumin. *Biosci. Biotechnol. Biochem.* 67, 1699–1705. doi: 10.1271/bbb.67.1699

Lending, C. R., and Larkins, B. A. (1992). Effect of the floury-2 locus on protein body formation during maize endosperm development. *Protoplasma* 171, 123–133. doi: 10.1007/BF01403727

Liu, H., Shi, J., Sun, C., Gong, H., Fan, X., Qiu, F., et al. (2016). Gene duplication confers enhanced expression of 27-kDa γ-zein for endosperm modification in quality protein maize. Proc. Natl. Acad. Sci. *U.S.A.* 113, 4964–4969. doi: 10.1073/pnas.1601352113

Mainieri, D., Rossi, M., Archinti, M., Bellucci, M., De Marchis, F., Vavassori, S., et al. (2004). Zeolin: A new recombinant storage protein constructed using maize γ-zein and bean phaseolin. *Plant Physiol.* 136, 3447–3456. doi: 10.1104/pp.104.046409

Martin, M. N., Tarczynski, M. C., Shen, B., and Leustek, T. (2005). The role of 5′-adenylylsulfate reductase in controlling sulfate reduction in plants. *Photosynth. Res.* 86, 309–323. doi: 10.1007/s11120-005-9006-z

Martinez, Y., Li, X., Liu, G., Bin, P., Yan, W., Mas, D., et al. (2017). The role of methionine on metabolism, oxidative stress, and diseases. *Amino Acids* 49, 2091–2098. doi: 10.1007/s00726-017-2494-2

Matityahu, I., Godo, I., Hacham, Y., and Amir, R. (2013). Tobacco seeds expressing feedback–insensitive cystathionine gamma–synthase exhibit elevated content of methionine and altered primary metabolic profile. *BMC Plant Biol.* 13, 206. doi: 10.1186/1471-2229-13-206

Mertz, E. T., Bates, L. S., and Nelson, O. E. (1964). Mutant gene that changes protein composition and increases lysine content of maize endosperm. *Sci.* 145, 279–280. doi: 10.1126/science.145.3629.279

Messing, J., and Fisher, H. (1991). Maternal effect on high methionine levels in hybrid corn. *J. Biotechnol.* 21, 229–237. doi: 10.1016/0168-1656(91)90044-V

Molvig, L., Tabe, L. M., Eggum, B. O., Moore, A. E., Craig, S., Spencer, D., et al. (1997). Enhanced methionine levels and increased nutritive value of seeds of transgenic lupins (*Lupinus angustifolius* L.) expressing a sunflower seed albumin gene. *Proc. Natl. Acad. Sci.* 94, 8393–8398. doi: 10.1073/pnas.94.16.8393

Mota-Martorell, N., Jove, M., Borras, C., Berdun, R., Obis, E., Sol, J., et al. (2021). Methionine transsulfuration pathway is upregulated in long-lived humans. *Free Radic. Biol. Med.* 162, 38–52. doi: 10.1016/j.freeradbiomed.2020.11.026

Navik, U., Sheth, V. G., Khurana, A., Jawalekar, S. S., Allawadhi, P., Gaddam, R. R., et al. (2021). Methionine as a double-edged sword in health and disease: Current perspective and future challenges. *Ageing Res. Rev.* 72, 101500. doi: 10.1016/j.arr.2021.101500

Nelson, O. E., Mertz, E. T., and Bates, L. S. (1965). Second mutant gene affecting the amino acid pattern of maize endosperm proteins. *Sci.* 150, 1469–1470. doi: 10.1126/science.150.3702.1469

Neubauer, C., and Landecker, H. (2021). A planetary health perspective on synthetic methionine. *Lancet Planet Health* 5, e560–e569. doi: 10.1016/S2542-5196(21)00138-8

Newell, M. A., Vogel, K. E., Adams, M., Aydin, N., Bodnar, A. L., Ali, M., et al. (2014). Genetic and biochemical differences in populations bred for extremes in maize grain methionine concentration. *BMC Plant Biol.* 14, 49. doi: 10.1186/1471-2229-14-49

Nguyen, H. C., Hoefgen, R., and Hesse, H. (2012). Improving the nutritive value of rice seeds: elevation of cysteine and methionine contents in rice plants by ectopic expression of a bacterial serine acetyltransferase. *J. Exp. Bot.* 63, 5991–6001. doi: 10.1093/jxb/ers253

Olsen, M. S., Krone, T. L., and Phillips, R. L. (2003). BSSS53 as a donor source for increased whole-kernel methionine in maize. *Crop Sci.* 43, 1634–1642. doi: 10.2135/cropsci2003.1634

Panda, A. K., Raju, M. V., Rao, S. V., Lavanya, G., Reddy, E., and Sunder, G. S. (2010). Replacement of normal maize with quality protein maize on performance, immune response and carcass characteristics of broiler chickens. *Asian-Australas. J. Anim. Sci.* 23, 1626–1631. doi: 10.5713/ajas.2010.10036

Pedernera, M., Celi, P., Garcia, S. C., Salvin, H. E., Barchia, I., and Fulkerson, W. J. (2010). Effect of diet, energy balance and milk production on oxidative stress in early-lactating dairy cows grazing pasture. *Vet. J.* 186, 352–357. doi: 10.1016/j.tvjl.2009.09.003

Phillips, R. L., and Mcclure, B. A. (1985). Elevated protein-bound methionine in seeds of a maize line resistant to lysine plus threonine. *Cereal Chem.* 62, 213–218.

Pi, T., Wei, S., Jiang, Y., and Shi, J. S. (2021). High methionine diet-induced Alzheimer's disease like symptoms are accompanied by 5-methylcytosine elevated levels in the brain. *Behav. Neurol.* 2021, 1–16. doi: 10.1155/2021/6683318

Planta, J., Xiang, X., Leustek, T., and Messing, J. (2017). Engineering sulfur storage in maize seed proteins without apparent yield loss. *Proc. Natl. Acad. Sci.* 114, 11386–11391. doi: 10.1073/pnas.1714805114

Radwanski, E. R., and Last, R. L. (1995). Tryptophan biosynthesis and metabolism: biochemical and molecular genetics. *Plant Cell.* 7, 921–934. doi: 10.1105/tpc.7.7.921

Ranocha, P., McNeil, S. D., Ziemak, M. J., Li, C., Tarczynski, M. C., and Hanson, A. D. (2001). The S-methylmethionine cycle in angiosperms: ubiquity, antiquity and activity. *Plant J.* 25, 575–584. doi: 10.1046/j.1365-313x.2001.00988.x

Robinson, J. L., and Bertolo, R. F. (2016). The pediatric methionine requirement should incorporate remethylation potential and transmethylation demands. *Adv. Nutr.* 7, 523–534. doi: 10.3945/an.115.010843

Robutti, J. L., Hoseney, R. C., and Deyoe, C. W. (1974). Modified opaque-2 corn endosperms. I. Protein distribution and amino acid composition. *Cereal Chem.* 51, 163–172.

Rojas, A., Betran, J. O., Scott, P., Atta, H., and Menz, M. (2010). Quantitative trait loci for endosperm modification and amino acid contents in quality protein maize. *Crop Sci.* 50, 870–879. doi: 10.2135/cropsci2008.10.0634

Ruan, T., Li, L., Peng, X., and Wu, B. (2017). Effects of methionine on the immune function in animals. *Health N. Hav.* 09, 857–869. doi: 10.4236/health.2017.95061

Saalbach, I., Waddell, D., Pickardt, T., Schieder, O., and Muntz, K. (1995). Stable expression of the Sulphur-rich 2S albumin gene in transgenic *Vicia narbonensis* increases the methionine content of seeds. *J. Plant Physiol.* 145, 674–681. doi: 10.1016/S0176-1617(11)81280-0

Sato, W., Furuta, C., Akomo, P., Bahwere, P., Collins, S., Sadler, K., et al. (2021). Amino acid-enriched plant-based RUTF treatment was not inferior to peanut-milk RUTF treatment in restoring plasma amino acid levels among patients with oedematous or non-oedematous malnutrition. *Sci. Rep.* 11, 12582. doi: 10.1038/s41598-021-91807-x

Schickler, H., Benner, M. S., and Messing, J. (1993). Repression of the high–methionine zein gene in the maize inbred line Mo17. *Plant J.* 3, 221–229. doi: 10.1046/j.1365-313X.1993.t01-14-00999.x

Scott, M. P., Bhatnagar, S., and Betran, J. (2004). Tryptophan and methionine levels in quality protein maize breeding germplasm. *Maydica* 49, 303–311.

Segal, G., Song, R., and Messing, J. (2003). A new opaque variant of maize by a single dominant rna-interference-inducing transgene. *Genetics* 165, 387–397. doi: 10.1093/genetics/165.1.387

Sethi, M., Singh, A., Kaur, H., Phagna, R. K., Rakshit, S., and Chaudhary, D. P. (2021). Expression profile of protein fractions in the developing kernel of normal, opaque-2 and quality protein maize. *Sci. Rep.* 11, 2469. doi: 10.1038/s41598-021-81906-0

Seymour, W. M. (2016). "Role of methionine and methionine precursors in transition cow nutrition with emphasis on liver function," in *2016, Florida ruminant nutrition symposium*, (Florida: University of Florida) 11–16.

Shen, B., Li, C., and Tarczynski, M. C. (2002). High free-methionine and decreased lignin content result from a mutation in the *Arabidopsis S*-adenosyl-L-methionine synthetase 3 gene. *Plant J.* 29, 371–380. doi: 10.1046/j.1365-313X.2002.01221.x

Siener, R., Struwe, F., and Hesse, A. (2016). Effect of L-methionine on the risk of phosphate stone formation. *Urology* 98, 39–43. doi: 10.1016/j.urology.2016.08.007

Singhal, N. K., Freeman, E., Arning, E., Wasek, B., Clements, R., Sheppard, C., et al. (2018). Dysregulation of methionine metabolism in multiple sclerosis. *Neurochem. Int.* 112, 1–4. doi: 10.1016/j.neuint.2017.10.011

Slater, A., Nigel, W. S., and Mark, R. F. (2008). *Plant biotechnology: The genetic manipulation of plants, 2nd eds* (India: Oxford University Press), 256–257.

Smet, K., Raes, K., Huyghebaert, G., Haak, L., Arnouts, S., and De Smet, S. (2008). Lipid and protein oxidation of broiler meat as influenced by dietary natural antioxidant supplementation. *Poult. Sci.* 87, 1682–1688. doi: 10.3382/ps.2007-00384

Soave, C., and Salamini, F. (1984). The role of structural and regulatory genes in the development of maize endosperm. *Dev. Genet.* 5, 1–25. doi: 10.1002/dvg.1020050102

Soladoye, O. P., Juarez, M. L., Aalhus, J. L., Shand, P., and Estevez, M. (2015). Protein oxidation in processed meat: Mechanisms and potential implications on human health. *Compr. Rev. Food Sci. Food Saf.* 14, 106–122. doi: 10.1111/1541-4337.12127

Song, R., Llaca, V., Linton, E., and Messing, J. (2001). Sequence, regulation, and evolution of the maize 22-kDa α zein gene family. *Genome Res.* 11, 1817–1825. doi: 10.1101/gr.197301

Song, R., and Messing, J. (2002). Contiguous genomic DNA sequence comprising the 19-kD zein gene family from maize. *Plant Physiol.* 130, 1626–1635. doi: 10.1104/pp.012179

Stadtman, E. R. (2001). Protein oxidation in aging and age-related diseases. *Ann. N. Y. Acad. Sci.* 928, 22–38. doi: 10.1111/j.1749-6632.2001.tb05632.x

Sultana, R., Perluigi, M., and Butterfield, D. A. (2009). Oxidatively modified proteins in Alzheimer's disease (AD), mild cognitive impairment and animal models of AD: Role of Abeta in pathogenesis. *Acta Neuropathol.* 118, 131–150. doi: 10.1007/s00401-009-0517-0

Swarup, S., Timmermans, M. C. P., Chaudhuri, S., and Messing, J. (1995). Determinants of the high-methionine trait in wild and exotic germplasm may have

escaped selection during early cultivation of maize. *Plant J.* 8, 359–368. doi: 10.1046/j.1365-313X.1995.08030359.x

Swigonova, Z., Lai, J., Ma, J., Ramakrishna, W., Llaca, V., Bennetzen, J. L., et al. (2004). Close split of sorghum and maize genome progenitors. *Genome Res.* 14, 1916–1923. doi: 10.1101/gr.2332504

Tabe, L. M., and Droux, M. (2001). Sulfur assimilation in developing Lupin cotyledons could contribute significantly to the accumulation of organic sulfur reserves in the seed. *Plant Physiol.* 126, 176–187. doi: 10.1104/pp.126.1.176

Tabe, L., Wirtz, M., Molvig, L., Droux, M., and Hell, R. (2010). Overexpression of serine acetlytransferase produced large increases in O-acetylserine and free cysteine in developing seeds of a grain legume. *J. Exp. Bot.* 61, 721–733. doi: 10.1093/jxb/erp338

Tan, Q., Zhang, L., Grant, J., Cooper, P., and Tegeder, M. (2010). Increased phloem transport of *S*-methylmethionine positively affects sulfur and nitrogen metabolism and seed development in pea plants. *Plant Physiol.* 154, 1886–1896. doi: 10.1104/pp.110.166389

Tapia-Rojas, C., Lindsay, C. B., Montecinos-Oliva, C., Arrazola, M. S., Retamales, R. M., Bunout, D., et al. (2015). Is L-methionine a trigger factor for Alzheimer's–like neurodegeneration?: Changes in Aβ oligomers, tau phosphorylation, synaptic proteins, Wnt signaling and behavioral impairment in wild-type mice. *Mol. Neurodegener.* 10, 62. doi: 10.1186/s13024-015-0057-0

Taylor, M., Chapman, R., Beyaert, R., Hernandez-Sebastia, C., and Marsolais, F. (2008). Seed storage protein deficiency improves sulfur amino acid content in common bean (*Phaseolus vulgaris* L.): Redirection of sulfur from γ-glutamyl-*S*-methyl-cysteine. *J. Agric. Food Chem.* 56, 5647–5654. doi: 10.1021/jf800787y

Toohey, J. I. (2014). Sulfur amino acids in diet–induced fatty liver: a new perspective based on recent findings. *Molecules* 19, 8334–8349. doi: 10.3391/molecules19068334

Ufaz, S., and Galili, G. (2008). Improving the content of essential amino acids in crop plants: Goals and opportunities. *Plant Physiol.* 147, 954–961. doi: 10.1104/pp.108.118091

Vasal, S. K. (2000). The quality protein maize story. *Food Nutr. Bull.* 21, 445–450. doi: 10.1177/156482650002100420

Wanders, D., Hobson, K., and Ji, X. (2020). Methionine restriction and cancer biology. *Nutrients* 12, 684. doi: 10.3390/nu12030684

Wang, X., and Larkins, B. A. (2001). Genetic analysis of amino acid accumulation in *opaque-2* maize endosperm. *Plant Physiol.* 125, 1766–1777. doi: 10.1104/pp.125.4.1766

Wang, X., Lopez-Valenzuela, J. A., Gibbon, B. C., Gakiere, B., Galili, G., and Larkins, B. A. (2007). Characterization of monofunctional aspartate kinase genes in maize and their relationship with free amino acid content in the endosperm. *J. Exp. Bot.* 58, 2653–2660. doi: 10.1093/jxb/erm100

Wang, L. Q., Zhong, M., Li, X. H., Yuan, D. J., Xu, Y. B., Liu, H. F., et al. (2008). The QTL controlling amino acid content in grains of rice (*Oryza sativa*) are co-localized with the regions involved in the amino acid metabolism pathway. *Mol. Breed.* 21, 127–137. doi: 10.1007/s11032-007-9141-7

Wen, C., Wu, P., Chen, Y., Wang, T., and Zhou, Y. (2014). Methionine improves the performance and breast muscle growth of broilers with lower hatching weight by altering the expression of genes associated with the insulin–like growth factor-I signalling pathway. *Br. J. Nutt.* 111 (2), 201–206. doi: 10.1017/S0007114513002419

Wenefrida, I., Utomo, H., Blanche, S., and Linscombe, S. (2009). Enhancing essential amino acids and health benefit components in grain crops for improved nutritional values. *Recent Pat. DNA Gene Seq.* 3, 219–225. doi: 10.2174/187221509789318405

Wheeler, K. B., and Latshaw, J. D. (1981). Sulfur amino acid requirements and interactions in broilers during two growth periods. *Poult. Sci.* 60, 228–236. doi: 10.3382/ps.0600228

Whitcomb, S. J., Nguyen, H. C., Bruckner, F., Hesse, H., and Hoefgen, R. (2018). Cystathionine gamma–synthase activity in rice is developmentally regulated and strongly correlated with sulfate. *Plant Sci.* 270, 234–244. doi: 10.1016/j.plantsci.2018.02.016

Wu, Y., Goettel, W., and Messing, J. (2009). Non-Mendelian regulation and allelic variation of methionine-rich delta-zein genes in maize. *Theor. Appl. Genet.* 119, 721–731. doi: 10.1007/s00122-009-1083-5

Wu, Y., and Messing, J. (2010). RNA interference–mediated change in protein body morphology and seed opacity through loss of different zein proteins. *Plant Physiol.* 153, 337–347. doi: 10.1104/pp.110.154690

Wu, Y., and Messing, J. (2014). Proteome balancing of the maize seed for higher nutritional value. *Front. Plant Sci.* 5. doi: 10.3389/fpls.2014.00240

Wu, Y., Wang, W., and Messing, J. (2012). Balancing of sulfur storage in maize seed. *BMC Plant Biol.* 12, 77. doi: 10.1186/1471-2229-12-77

Xiang, X., Hu, B., Pu, Z., Wang, L., Leustek, T., and Li, C. (2022). Co-overexpression of *AtSAT1* and *EcPAPR* improves seed nutritional value in maize. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.969763

Xiang, X., Wu, Y., Planta, J., Messing, J., and Leustek, T. (2018). Overexpression of serine acetyltransferase in maize leaves increases seed-specific methionine-rich zeins. *Plant Biotechnol. J.* 16, 1057–1067. doi: 10.1111/pbi.12851

Xiong, Y. L., and Guo, A. (2020). Animal and plant protein oxidation: chemical and functional property significance. *Foods* 10 (1), 40. doi: 10.3390/foods10010040

Xu, J. H., and Messing, J. (2008). Organization of the prolamin gene family provides insight into the evolution of the maize genome and gene duplications in grass species. *Proc. Natl. Acad. Sci.* 105, 14330–14335. doi: 10.1073/pnas.0807026105

Yin, F., Pajak, A., Chapman, R., Sharpe, A., Huang, S., and Marsolais, F. (2011). Analysis of common bean expressed sequence tags identifies sulfur metabolic pathways active in seed and sulfur–rich proteins highly expressed in the absence of phaseolin and major lectins. *BMC Genom.* 12, 268. doi: 10.1186/1471-2164-12-268

Yoo, S. C. (2017). Quantitative trait loci controlling the amino acid content in rice (*Oryza sativa* L.). *J. Plant Biotechnol.* 44, 349–355. doi: 10.5010/JPB.2017.44.4.349

Yu, D., Yang, S. E., Miller, B. R., Wisinski, J. A., Sherman, D. S., Brinkman, J. A., et al. (2018). Short-term methionine deprivation improves metabolic health via sexually dimorphic, mTORCI-independent mechanisms. *FASEB. J.* 32 (6), 3471–3482. doi: 10.1096/fj.201701211R

Zeh, M., Casazza, A. P., Kreft, O., Roessner, U., Bieberich, K., Willmitzer, L., et al. (2001). Antisense inhibition of threonine synthase leads to high methionine content in transgenic potato plants. *Plant Physiol.* 127, 792–802. doi: 10.1104/pp.010438

Zhang, Y., Schernthaner, J., Labbe, N., Hefford, M. A., Zhao, J., and Simmonds, D. H. (2014). Improved protein quality in transgenic soybean expressing a *de novo* synthetic protein, MB-16. *Transgenic Res.* 23, 455–467. doi: 10.1007/s11248–013–9777–5

Zhong, M., Wang, L., Yuan, D., Luo, L., Xu, C., and He, Y. (2011). Identification of QTL affecting protein and amino acid contents in rice. *Rice Sci.* 18, 187–195. doi: 10.1016/S1672-6308(11)60026-7

Zhou, Y., and Staatz, J. (2016). Projected demand and supply for various foods in West Africa: Implications for investments and food policy. *Food Policy* 61, 198–212. doi: 10.1016/j.foodpol.2016.04.002

frontiers | Frontiers in Plant Science

TYPE Review
PUBLISHED 27 October 2023
DOI 10.3389/fpls.2023.1231013

OPEN ACCESS

EDITED BY
Baohong Zhang,
East Carolina University, United States

REVIEWED BY
Mohd Fadhli Hamdan,
University of Malaya, Malaysia
Siddanna Savadi,
Directorate of Cashew Research (ICAR), India

*CORRESPONDENCE
Dragana Miladinović
dragana.miladinovic@ifvcns.ns.ac.rs

RECEIVED 29 May 2023
ACCEPTED 09 October 2023
PUBLISHED 27 October 2023

CITATION
Yıldırım K, Miladinović D, Sweet J, Akin M, Galović V, Kavas M, Zlatković M and de Andrade E (2023) Genome editing for healthy crops: traits, tools and impacts. *Front. Plant Sci.* 14:1231013. doi: 10.3389/fpls.2023.1231013

Genome editing for healthy crops: traits, tools and impacts

Kubilay Yıldırım[1], Dragana Miladinović[2]*, Jeremy Sweet[3], Meleksen Akin[4], Vladislava Galović[5], Musa Kavas[6], Milica Zlatković[5] and Eugenia de Andrade[7,8]

[1]Department of Molecular Biology and Genetics, Faculty of Arts and Sciences, Ondokuz Mayıs University, Samsun, Türkiye, [2]Institute of Field and Vegetable Crops, National Institute of Republic of Serbia, Novi Sad, Serbia, [3]Sweet Environmental Consultants, Cambridge, United Kingdom, [4]Department of Horticulture, Iğdır University, Iğdır, Türkiye, [5]Institute of Lowland Forestry and Environment (ILFE), University of Novi Sad, Novi Sad, Serbia, [6]Department of Agricultural Biotechnology, Faculty of Agriculture, Ondokuz Mayıs University, Samsun, Türkiye, [7]National Institute for Agricultural and Veterinary Research (INIAV), I.P., Oeiras, Portugal, [8]GREEN-IT Bioresources for Sustainability, ITQB NOVA, Oeiras, Portugal

Crop cultivars in commercial use have often been selected because they show high levels of resistance to pathogens. However, widespread cultivation of these crops for many years in the environments favorable to a pathogen requires durable forms of resistance to maintain "healthy crops". Breeding of new varieties tolerant/resistant to biotic stresses by incorporating genetic components related to durable resistance, developing new breeding methods and new active molecules, and improving the Integrated Pest Management strategies have been of great value, but their effectiveness is being challenged by the newly emerging diseases and the rapid change of pathogens due to climatic changes. Genome editing has provided new tools and methods to characterize defense-related genes in crops and improve crop resilience to disease pathogens providing improved food security and future sustainable agricultural systems. In this review, we discuss the principal traits, tools and impacts of utilizing genome editing techniques for achieving of durable resilience and a "healthy plants" concept.

KEYWORDS

CRISPR, crops, crop improvement, pathogens, resilience, durable resistance, fungal, bacterial and virus infections, parasitic weeds

1 Introduction

Crops are grown in different geographic, climatic, and agricultural conditions, where they are challenged by a vast range of pests and diseases that can substantially reduce crop yields and production (Savary et al., 2019). Managing these biotic stresses usually involves considerable effort and expense for farmers, particularly when these stressors have the ability to adapt to certain control measures. Chemical pesticides provide a level of protection, but often reliance on them is unsustainable due to resistance development, and environmental concerns (Lykogianni et al., 2021). Reduction of the efficacy of

pesticides due to rapid pathogen evolution and resistance development by adaptation under selection pressure has been extensively documented for chemical pesticides (McDonald, 2014; Akın et al., 2023). Although the use of antagonistic microorganisms for biological control has advanced significantly, there are still few approved biofungicides in the market due to issues with their effectiveness, legislation, and registration procedures (Collinge et al., 2022). An efficient and alternative method to protect crops from pests and diseases is the cultivation of resistant plant genotypes in agriculture (Gvozdenac et al., 2022; Kavas et al., 2023).

Viruses, bacteria, filamentous pathogens (fungi and oomycetes) and parasitic weeds are the major groups of plant pathogens that can affect crops both in the field and post-harvest (Strange and Scott, 2005). The effects of these biotic threats on agricultural production range from none or mild symptoms to pandemics that seriously compromise crop production over large cultivation areas. Plant pathogens can be introduced into new areas through various means, such as contaminated plant material, infected seeds, soil, or infected tools and equipment. International trade and transportation of agricultural products can also facilitate the movement of pathogens across regions. (Bisht et al., 2019). Understanding the specific characteristics and modes of transmission of a particular plant pathogen is essential for developing effective bredding strategies to prevent its spread and manage diseases in agricultural and natural settings. Integrated pest management (IPM) approaches that combine with plant defense mechanisms are often used to mitigate the impact of plant pathogens and minimize their spread (Kocmánková et al., 2009).

'Healthy plants' are vital to sustainable and profitable crop production and to the quality and cost of the nation's supply of food, fuel, and fiber. However, maintaining "healthy plants" is a challenge due to climate and other environmental changes that can disrupt the interactions between species (Tamura et al., 2022) in a range of environments (Karavolias et al., 2021). Currently, climate change is favoring enlargement of the geographical distribution of some already existing and newly emerging pests and invasive plants (Jones and Barbetti, 2012; King et al., 2018). Furthermore, the markets and economy require extensive movement of plants and agricultural goods between continents, facilitating the movement of pathogens, along with human activities such as travel and urbanization that promote the entry of new pathogens into agricultural ecosystems compromising crop health (Franic et al., 2022).

Crop cultivars in commercial use have often been selected because they show high levels of resistance to pathogens. However, widespread cultivation of these crops for many years in the environments favorable to a pathogen requires durable forms of resistance to maintain "healthy plants". This durable resistance depends on the variability of pathogenicity, and the nature of the resistance mechanisms in crop cultivars (Nnadi and Carter, 2021). Many pathogens are heterozygous many different pathotypes or races which can rapidly adapt to new environments or hosts. Some pathogens are host-specific, whereas others have a diversity of hosts and thus can maintain reservoirs of infective pathotypes with a greater ability to evolve and adapt to climate and weather conditions (Amari et al., 2021).

Breeding of new varieties tolerant/resistant to biotic stresses by incorporating genetic components related to durable resistance, developing new breeding methods and new active molecules, and improving the IPM strategies have been of great value, but their effectiveness is challenged by the newly emerging diseases and the rapid change of pathogens due to climatic changes (Hussain, 2015; Bhoi et al., 2022). Achieving continuous production of resistant varieties needs continuous adjustment of breeding methods. Hence, in recent years, novel methods to enhance genetic resistance have been developed. These include changing the genetics of crop plants, introducing novel genes into plants and the expression of interfering RNAs (RNAi). Developing disease-resistant crops through genome editing-based techniques offers an effective, environmentally friendly, low-input, and sustainable approach to plant disease management (Ali et al., 2022). Their effective application has been supported by the characterization of many immune receptors, ("*R*" genes) and the genetic basis of cell surface immunity in the last decades. This progress has enabled us to understand the molecular basis of interactions between plants and pathogens and plant innate immunity, both of which are essential for developing disease-resistant plant varieties (Ali et al., 2022; Bhoi et al., 2022).

Genome editing (GE) technology has developed new tools and methods to identify genes involved in defence in crops and to increase crop resilience to pathogens thus providing improved food security within agricultural systems that are more sustainable (Gosavi et al., 2020). Here we review the main traits, tools and impacts of the application of genome editing techniques in crop improvement for the achievement of durable resilience and a "healthy crops" concept.

2 Cross-talk between plants/ pathogens via plant immunity

Plants and pathogens have an endless complex co-evolutionary arms race where pathogens try to overcome plant defenses and in turn, plants have developed a range of defence mechanisms to detect and prevent pathogen invasion ("zig-zag model") (Jones and Dangl, 2006). Throughout evolution, plants have been armed with several physical barriers and biochemical adaptations to prevent the entry of pathogens into the plant cells. In addition, the plant immune system has been developed according to the complexity of the feeding behaviors of pathogens through co-evolution over millions of years (Voigt, 2014). Plant pathogens can be biotrophs that completely or partially rely on host cells for the completion of their life cycle. These types of pathogens manipulate the host metabolism to induce favorable nutritional conditions and maintain host viability to acquire nutrients as much as possible. Biographies cause relatively minor damage to the host plant cell, while necrotrophic pathogens kill their hosts during infection by using all their sources. Plants do not have an adaptive immune system due to their lack of specialized immune cells. Nevertheless, plants developed resistance to biotrophs and necrotrophs with induced signal transduction routes that share cross-talk and

independent pathways. This plant's innate immune system is based on pathogen receptors detecting the presence of pathogens (immune recognition) and molecular signalling pathways to transmit the message of invasion (signal integration) to the cell nucleus (Andolfo and Ercolano, 2015). Signal integration of invasion alters the transcriptional gene expression in the nucleus and activates the defence response in host plant cell (Figure 1).

2.1 Plant innate immunity

If a pathogen manages to enter a host, a multi-layered innate immune system is activated as a defense response (Jones and Dangl, 2006; Andolfo and Ercolano, 2015). The first layer of the defense comprises receptor-like proteins or receptor-like protein kinases known as plant/pathogen recognition receptors (PRRs) that detect pathogens at the plant cell membrane surface and in the apoplast (Wise et al., 2007; Dodds and Rathjen, 2010). At the plant cell membrane, PRRs "recognize" conserved microbial elicitors known as microbe-associated molecular patterns (MAMPs, also referred to as pathogen-associated molecular patterns-PAMPs), and pathogen proteins (apoplastic effectors) that are produced in the apoplast and this initiates a plant defence response called MAMP/PAMP-triggered immunity (MTI/PTI) (Jones and Dangl, 2006; Thomma et al., 2011; Andolfo and Ercolano, 2015; Boschi et al., 2017; Boutrot and Zipfel, 2017). Moreover, the pathogen attack can trigger plant signals called Damage-Associated Molecular Patterns (DAMPs) that can also activate PTI (Hou et al., 2019). The MAMPs/

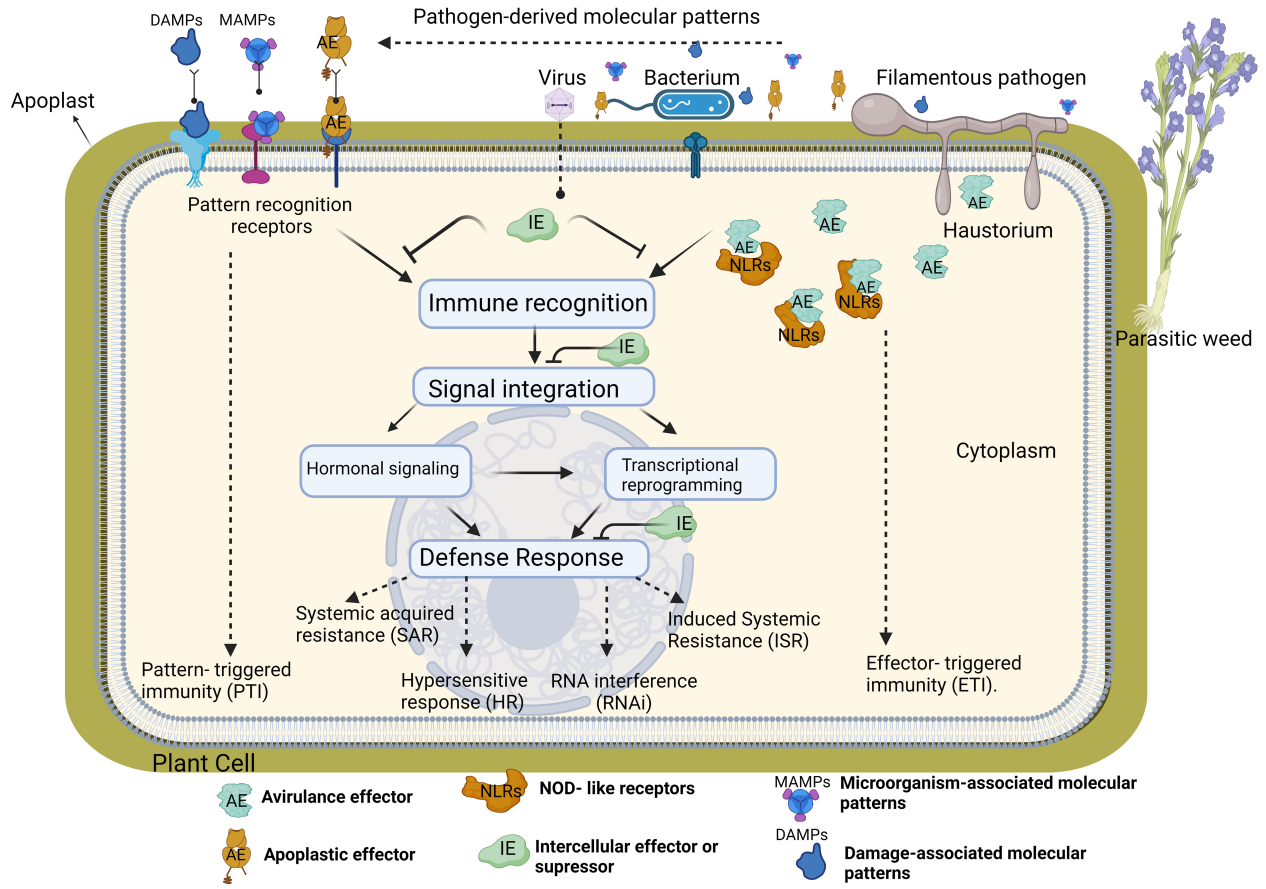


FIGURE 1
The activation of the plant innate immune system requires tree steps; immune recognition, signal integration and defense response. Plants use numerous cell surface and intracellular immune receptors to recognize microorganism/host-derived molecular patterns (MAMPs and DAMPs), or apoplastic/avirulance effectors (AE). Cell surface pattern recognition receptors (PRRs) bind to MAMPs or DAMPs or AE directly through their extracellular domain while NOD- like receptors (NLRs) recognize effectors delivered inside host cells by directly binding effectors or sensing modulation of effector host targets. PRR- mediated recognition of MAMPs or DAMPs elicits pattern- triggered immunity (PTI), and NLR- mediated pathways trigger effector-triggered immunity (ETI). Activation of immune receptors subsequently initiates the second phase of immune system. In this phase, various immune signaling events such as calcium fluxes, activation of mitogen- activated protein kinase (MAPK) cascades, alteration of host transcription and phytohormone signaling trigger the defense response in each cellular compartment in plants. Hormone- dependent response generally activates a large set of plant defense-related genes against biotrophs. For instance, hormone accumulation in plants triggers hypersensitive response (HR) which cause rapid local death of the infected and surrounding cells to restrict the spread of pathogens to other parts of the plant. Accumulation of hormones and pathogenesis-related proteins in the plants can also induce long-lasting protection against a broad spectrum of pathogens, called systemic acquired resistance (SAR). In this resistance, putative SAR signal molecules move from the infected systemic organs to non-infected distant parts of the plant where it make more resistance to pathogens prior to infection. Induced Systemic Resistance (ISR) is another defense response increasing physical or chemical barriers of the host plant against pathogens rather than direct killing or inhibiting the invading pathogen. RNA interference (RNAi) is the last plant resistance mechanism activated during the viral infection.

PAMPs include several components: bacterial flagellin, elongation factor thermo-unstable (EF-Tu), and fungal chitin, whereas DAMPs are molecules that are released from damaged cells undergoing pathogen invasion (Figure 1) (Lanna-Filho, 2023). As the battle continues, the plant produces reactive oxygen species (ROS) and secrets antimicrobial products such as phytoalexins, and phenolic compounds like flavonoids and tannins in the intercellular spaces that can destroy pathogens (Doehlemann et al., 2008; Saijo et al., 2018; Kebert et al., 2022)

In the cytoplasm, the pathogen secret proteins (cytoplasmic effectors, formerly known as avirulence factors) that target plant susceptibility (*S*) genes to manipulate plant processes to support pathogen growth, promote disease development and induce susceptibility. This phenomenon is called effector-triggered susceptibility (ETS) (Jones and Dangl, 2006; Jones and Barbetti, 2012; Weßling et al., 2014; Hui et al., 2019). As a counter defence strategy, proteins encoded by disease resistance genes (*R* genes) recognize pathogen effectors (Zhang X. et al., 2017; Collinge, 2020). Most R proteins contain domain-rich amino acid leucine (leucine-rich repeat - LRR), and have a nucleotide binding site (NBS) and NOD-like receptors (NLRs) (Figure 1). The recognition of the pathogen effectors by plant *R* genes initiates NLR-mediated response known as NLR or effector-triggered immunity (NTI/ETI) to stop pathogen growth and development (Jones and Dangl, 2006; Dodds and Rathjen, 2010; Win et al., 2012; Lo Presti et al., 2015., Jones et al., 2016). This plant immunity response is generally stronger than pathogen-triggered immunity (PTI) (Jones and Dangl, 2006). ETI results in events like cell wall modifications (e.g. depositions of lignin and callose), stomata closure, expression of pathogenesis-related genes that induce production of proteins that show antimicrobial activity (e.g., chitinases, β 1-3 glucanases, defensins, peroxidases), secondary metabolites like phytoalexins and the accumulation of plant hormones related to plant defence, including salicylic acid (SA), jasmonic acid (JA) and ethylene (ET) (Mukhtar, 2013; Uehling et al., 2017; Andersen et al., 2018). The most extreme consequences of ETI include a hypersensitive response (HR) along with the generation of ROS that leads to programmed cell death (PCD) and the formation of necrotic lesions, where the infected plant cells kill themselves to protect other cells and restrict the spread of the pathogen from the infection site to neighboring cells (Figure 1) (Gong et al., 2019). Moreover, recently, Khattab et al. (2023) identified trans-ferulic acid, a monolignol precursor as a "plant surrender signal" that accumulates in grapevines under stress. The ferulic acid activates the secretion of the fungal phytotoxin fusicoccin A aglycone which stimulates programmed cell death after infection with the pathogenic necrotrophic fungus *Neofusicoccum parvum*.

2.2 RNAi and R gene-mediated plant immunity

Two key components have been described for plant-virus interactions and plant defence responses to viral pathogens; RNA silencing and R gene-mediated pathways. RNA gene silencing [also called RNA interference (RNAi)] is the main plant defence response to viral pathogens (Moon and Park, 2016). Most plant viruses have RNA genomes that contain a regulatory stem-loop. These loops are recognized by virus-encoded RNA-dependent RNA polymerases to copy the viral genome into complementary double-stranded RNAs (dsRNAs) (Ruiz-Ferrer and Voinnet, 2009). Host ribonuclease III-like protein, also called Dicer-like (DCL), recognizes the dsRNAs and then breaks them up into short interfering RNAs (siRNAs). The siRNAs (20-25 bp in length) have complementary sequences to the viruses and act as guides to direct RNA-induced silencing complex (RISC) in their target and degrade the viral RNA molecules (Mallory et al., 2008; Ruiz-Ferrer and Voinnet, 2009). Interestingly, plant viruses often encode viral suppressor RNAi (VSRs) to inactivate the plant RNAi-mediated silencing pathway and enhance viral replication, assembly, or movement (Ding and Voinnet, 2007). VSR-mediated suppression of antiviral RNA silencing pathway is known to occur in two ways. VSRs can sequester the small RNA duplexes to block their binding to viral dsRNAs (Lakatos et al., 2006) or directly impede the activity of RISC proteins to impair the assembly of the complex (Carbonell and Carrington, 2015). Besides RNAi, plants have also developed a second layer of dominant and recessive defence against viruses via resistance genes (*R*-genes) (De Ronde et al., 2014). Most of these *R*-genes are triggered by a virus and confer dominant resistance like in Ty-1 *R*-gene from tomato against tomato yellow leaf curl virus (TYLCV). This gene encodes an RNA-dependent RNA polymerase and confers resistance against TYLCV by amplifying the RNAi signal (Verlaan et al., 2013). Since viruses require host factors for their infection, cross-talk between such plant susceptibility factors and the virus may also lead to resistance (Moon and Park, 2016). For instance, some viruses encode a cap-like structure to interact with the host translation initiation factors (eIF4E/eIF4G) for the expression of the viral genome. Loss of function in these factors leads to a recessive resistance in plants (Truniger and Aranda, 2009). Indeed, viral pathogens generally encode proteins for the suppression of plant' RNAi defence mechanisms (Wang et al., 2012). Therefore, both RNAi and *R* gene-mediated pathways in plants undergo crosstalk to maximize the efficiency of defence responses against viral infections (Nakahara and Masuta, 2014). For example, in the *Arabidopsis* hypersensitive response to the turnip crinkle virus, the *HRT* genes respond to the TCV coat protein by producing a DNA-binding protein. HRT-mediated resistance requires double-stranded RNA-binding protein-4 which is also the component of the RNAi (Zhu et al., 2013) PTI also limits virus infection in plants and this defence response is mediated by dsRNA (Niehl and Heinlein, 2019).

2.3 Hormone-mediated immunity and crosstalk between plants and pathogens

Activation of PTI and ETI in infected tissues often triggers a third layer of plant immunity referred to as induced resistance (IR) and can occur at the site of the attack, in parts of plants distal from the site of infection, or throughout the entire plant (Figure 1). During those systemic immune responses, hormonal interactions and their signalling pathways play the role of central regulators in

plant defence against a wide range of pathogens and insects (Berens et al., 2017). Different hormones accumulate in plant tissues depending on the type of attacker and each hormone regulates its own immune network. Salicylic acid (SA) and Jasmonic acid (JA) are the two basic hormones forming the backbone of plant immune systems against pathogens and insects (Wasternack and Song, 2017; Zhang and Li, 2019). The SA-dependent response generally activates a large set of plant defence-related genes against biotrophs (Vos et al., 2015). For instance, SA accumulation in plants triggers a rapid local death of the infected and surrounding cells to restrict the spread of pathogens to other parts of the plant (Balint-Kurti, 2019). In addition to this rapid hypersensitive response, the accumulation of SA and pathogenesis-related proteins in plants can also induce long-lasting protection against a broad spectrum of microorganisms and insects, called systemic acquired resistance (SAR) (Backer et al., 2019). In this resistance, putative SAR signal molecules such as methyl salicylate move from the infected systemic organs to non-infected distant parts of the plant where it induces pathogenesis-related genes against pathogens. In this way, distant leaves or tissues become more resistant to pathogens before infection (Backer et al., 2019; Balint-Kurti, 2019). Induced Systemic Resistance (ISR) is another resistance strategy in plants that is activated by infection. This strategy depends on increasing the physical or chemical barriers of the host plant against pathogens rather than directly killing or inhibiting the invading pathogen. Plants are sensitized to produce an enhanced ISR response by infection with beneficial bacteria and fungi living in the rhizosphere and signal transduction pathways activated by JA (Yu et al., 2022). These root-associated mutualistic microbes boost plant defenses, rendering the entire plant more resistant to pathogens and pests (Backer et al., 2019; Yu et al., 2022).

JA and its oxylipin derivatives (jasmonates) are generally synthesized and accumulated in plants in response to herbivore arthropods or infection with necrotrophs (Wang et al., 2021). Some herbivore insects take their nutrients from plants by mechanical damage of plant tissues while necrotrophs derive their energy from dead or dying cells (Vega-Muñoz et al., 2020). During the insect chewing or wounding during herbivory, JA is rapidly synthesized locally on the damaged part of the plant and systemically in parts of plants not affected by pathogens (Wang et al., 2021). This increase in JA concentration activates the expression of defense-related genes that induce production of toxic secondary metabolites, formation of physical barrier (such as trichome) and generation of volatile organic compounds (VOCs) (Escobar-Bravo et al., 2017; Vega-Muñoz et al., 2020). However, some biotrophic pathogens and hemibiotrophic pathogens develop mechanisms to evade this JA-mediated plant defense through injecting toxins and virulence-effector proteins into host cells to suppress JA signaling components (Vargas et al., 2012; Wang et al., 2021). SA and JA can act alone or show synergistic and antagonistic interactions with each other or with other hormones in a complex interplay (Liu et al., 2016). This phenomenon is known as hormone crosstalk and is an important component of the architecture of the plant immune signaling network (Yang et al., 2019). For instance, JA pathway is divided into two branches (Pieterse et al., 2014; Yıldırım and Kaya, 2017). The ERF branch of the JA pathway is co-regulated by ethylene (ET). This branch is activated by infection with necrotrophic pathogens. The second branch of JA pathway, MYC branch, is co-regulated by abscisic acid (ABA) to provide protections against chewing insects (Aerts et al., 2021). It has also been shown that JA signaling can block SA accumulation in plants through modulation of multiple transcription factors (Caarls et al., 2015). This crosstalk between JA and SA signaling pathways has been reported to coordinately regulate plant disease resistance against necrotrophic or hemibiotrophic pathogens (Yang et al., 2015). SA is generally known to activate the expression of early defense-related genes, while JA induces late defense-related gene expression in infected plants (Caarls et al., 2015; Yang et al., 2015; Sucu et al., 2018; Aerts et al., 2021).

3 Tools for crop genome editing and introduction of durable resilience

Because plants and their pathogens have been evolving together, they have developed a sophisticated mode of communication where changes in virulence of the pathogen is being balanced by the changes in the resistance of the host, and vice versa. This plant-pathogen balance is known as "gene-for-gene concept" and it is an integral part of the plant's and pathogen's life cycle. The concept in which a single gene of the host corresponds to the single gene of the pathogen has proven extremely important in plant breeding (Hammond-Kosack and Jones, 1997; Zaidi et al., 2018; Naidoo et al., 2019; Li W. et al., 2020; Rato et al., 2021). However, it is a quite complex interaction since both plants and pathogens can have multiple genes that can affect their resistance and virulence, respectively and there are many different races of a single pathogen species that can infect different plant cultivars depending on the combination of their resistance genes. Pathogen strains that can induce resistance reaction in a plant have evolved dominant avirulence (Avr) genes, and as counter defence strategy plants have evolved dominant resistance (R) genes. In contrast, pathogen strains that can "sneak" by the plant's defensive system undetected have recessive virulence genes and these strains can cause the disease. The weakness of R-mediated resistance leads to the emergence of resistant pathogen strains and thus it is short-lived in the field (Zaidi et al., 2018; Li M. Y. et al., 2020; Tyagi et al., 2020; Pan et al., 2021; Rato et al., 2021). Moreover, this type of resistance is associated mainly with biotrophic and hemibiotrophic pathogens, whereas it is challenging to use these resistant strategies against necrotrophic pathogens due to their need to colonies the dead tissue (Collinge and Sarrocco, 2022).

Currently, several crop plants have fully sequenced genomes and these annotated genomes result in increased knowledge of the molecular details and genetic functions of plant genes. This knowledge is exploited by genome editing (GE) innovations creating a greater advancement in understanding the gene regulatory functions in plants, pathogens and their interactions. GE has been accepted as a new breeding technique and has been used to improve plant resistance against many kinds of pathogens in past decade (Secgin et al., 2022). GE techniques such as zinc-finger nucleases (ZFN), transcription-activator-like effector

nucleases (TALEN) use DNA nucleases guided with the engineered proteins. On the other hand, newly discovered CRISPR/Cas system depend on oligo-directed mutagenesis with sequence-specific nucleases. Due to its high accuracy, cost-effectiveness, and simplicity, the CRISPR/Cas9 system became the most popular GE tool for plant breeding (Aksoy et al., 2022). This system consists of the Cas protein inducing a double-strand break (DSB) in the DNA, and the single guide RNA (sgRNA) directing the Cas protein to the genomic target. The specificity of the system is conferred by easily programmable 20-nt-long guide RNA sequences complementary to the target genomic sequence (Cardi et al., 2023). CRISPR/Cas-mediated DSB can result in insertions or deletions (InDels) in the target DNA when repaired by the error-prone non-homologous end joining (NHEJ) mechanism. This would result in a simple random mutation in the target gene, most likely leading to a frameshift causing a loss-of-function phenotype. (Figure 2). CRISPR/Cas could be also used for the introduction of a sequence of choice via homology-directed repair (HDR) with the presence of a repair template in the complementary flanking arms. Such editing of the original gene sequence by introducing specific mutations can also be used to alter a single nucleotide in the genome to change the amino acid structure of the proteins, enzyme activities or substrate specificity [(Figure 2) Miladinović et al., 2021]. In recent years, dead or deactivated Cas (dCas)-based technologies have been developed and used for alteration of gene expression in plants. One of this technology is known as CRISPR activation or CRISPRa in which a catalytically dead (d) Cas9 is fused with a transcriptional effector to modulate target gene expression. Once the guide RNA navigates to the genome locus along with the effector arm, the dCas9 is unable to cut, and instead, the effector activates the downstream gene expression. On the contrary, CRISPR interference or CRISPRi technology just contains a catalytically dead (d) Cas9 and when guide RNA navigates to the genome locus along with the effector arm, it represses the downstream gene expression instead of

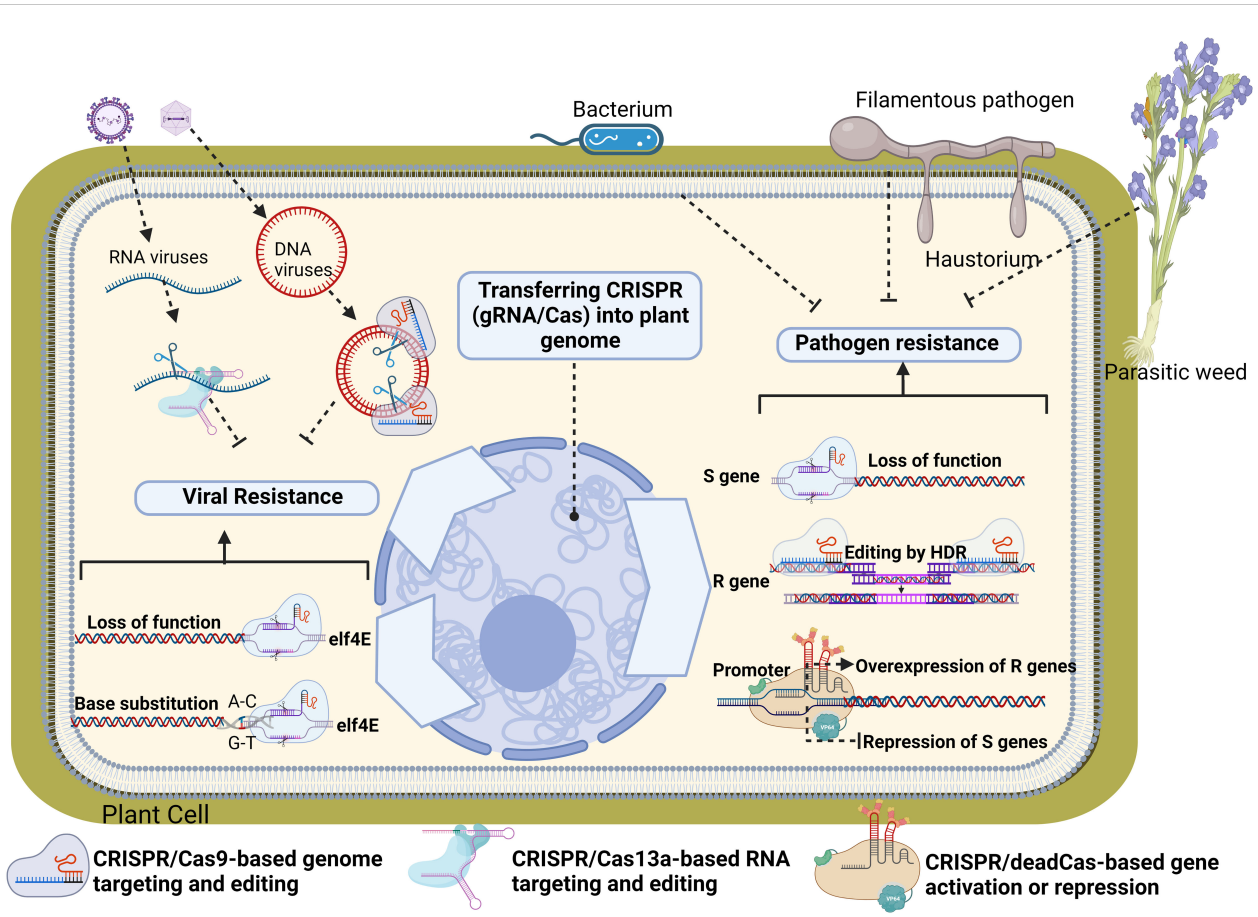


FIGURE 2
Theoretical and already tested CRISPR/Cas applications to increase plant resistance toward pathogens. CRISPR/Cas9 can be used to disrupt plant susceptibility (S) genes (such as Eukaryotic translation initiation factor 4E (elf4E)) by targeting coding regions to knock out these genes, or to alter sequences of promoter regions, precluding pathogen effector binding to the promoter and thus disrupting plant susceptibility. In addition, Dead Cas9-based CRISPR systems could be used to overexpression of resistance genes or suppression of S genes. CRISPR-mediated homology-directed repair (HDR) can be used to introduce resistance (R) genes against pathogens in cases where the plant-pathogen interaction (and S genes) is not well studied. To develop pathogen resistance without disrupting or replacing whole genes, CRISPR based base-edition technology can be used to achieve specific mutations (biomimicking) in genes to turn them into resistant genes against pathogens of interest. The native function of CRISPR can be also mimicked directly to target and interfere with the genomes of pathogens of interest without affecting plant genome. For example, CRISPR can interfere with DNA genomes of viruses through DNA-targeting gRNA/Cas9 systems or it can disrupt pathogen's RNA genomes through RNA-targeting gRNA/Cas13a systems. Loss of function in S genes.

activating it. In this section already tested genome editing approaches used to increase plant resistance toward pathogens will be listed and summarized with some theoretical applications.

3.1 Genome editing for viral resistance

Plant DNA and RNA virus families cause diseases and crop losses in a broad range of important crops. The dsDNA nature of DNA viruses, *Geminiviridae* and *Caulimoviridae*, make them good targets for CRISPR/Cas and this has become a popular approach to antiviral engineering in crops (Table 1). CRISPR-mediated resistance against DNA viruses was developed for *Cauliflower mosaic virus* (Liu et al., 2018), *Cotton leaf curl Multan virus* (Yin et al., 2019), Tomato yellow leaf curl virus (TYLCV), (Ali et al., 2015a; Ali et al., 2015b; Seçgin et al., 2021) *Beet severe curly top virus* (Baltes et al., 2015) and *Bean yellow dwarf virus* (Ji et al., 2015) in *Arabidopsis* and tobacco plants. gRNA/Cas9 constructs were designed to target and cleave viral replication (REP), coat protein and noncoding stem-loop sequences [TAATATTAC] common to all geminiviruses. Transient and stable expression of these constructs in transgenic plants exhibited high levels of viral resistance with significant reductions in virus accumulation and disease symptoms and revealed that the strongest virus inhibitory effect was achieved by the gRNA targeting the stem-loop sequence (Ali et al., 2015a; Baltes et al., 2015; Ji et al., 2015; Yin et al., 2019). This indicated that the stem-loop region could be a good target in CRISPR/Cas9-mediated resistance for broad-spectrum resistance to other geminiviruses. Other findings by Ali et al. (2015b) corresponded well with this suggestion that transient expression of gRNA/Cas9 construct confers resistance against mixed infection with Beet curly top virus and *Merremia mosaic virus* (MeMV), both of which share this conserved stem-loop sequence. In another approach, catalytically inactivated Cas9 (deadCas9) was successfully targeted to conserved stem-loop sequence of *Cotton leaf curl virus* to inhibit its replication and accumulation (Khan et al., 2019). In addition to model plants, CRISPR-mediated resistance against DNA viruses has also been carried out on sugar beet infections with *Beet Curly Top Iran Virus* (Yıldırım et al., 2022; Yıldırım et al., 2023), barley plants infected with wheat dwarf virus (Kis et al., 2019) and tomato infected with *TYLCV* (Tashkandi et al., 2018).

All these studies indicated the successful use of CRISPR/Cas9 to enhance virus resistance in plants. However, GE-based viral resistance in plants has some limitations. For example, targeting and mutating the virus genome could create a new variant of the virus that could be more aggressive and resistant to plant defence systems. Therefore, CRISPR systems targeting the multiple promoters or gene structures need to be designed to hinder mutant viral escape and to obtain full viral resistance. In addition, the requirement of PAM and dsDNA structure for effective digestion makes it impossible to target ssDNA structure of the viruses by CRISPR (Yıldırım et al., 2023). Fortunately, the newly discovered CRISPR/Cas systems offer precise and simple solutions to these problems (preventing viral escape, multiplexing DNA targeting, and even easy viral diagnostics). For instance, CRISPR/Cas12 has been realized to be much more versatile than CRISPR/Cas9 (Ali and Mahfouz, 2021). Cas12 requires only a short crRNA (making engineering easy), can process polycistronic crRNAs (making multiplexing possible, with no chance of virus escape, allowing multiple genomic loci to be edited at once), targets ssDNAs, and dsDNAs, and degrades ssDNAs via trans activity, Cas12 is comparatively small, and can easily be delivered via deconstructed viral vectors.

Furthermore, the nonspecific degradation of ssDNAs or ssRNAs (reporters) upon recognition of a specific target by Cas12, Cas13, and Cas14 variants provides the opportunity to develop an efficient diagnostic system for deployment in the field. Coupling of the target specificity and nuclease activity of Cas variants with target enrichment (via isothermal amplification, LAMP, RPA) and signal amplification (CONAN or SENSR) has the potential to change the entire scope of plant virus detection and control measures. Discovery of RNA-targeting Cas endonucleases (FnCas9 and Cas13) offered new possibilities for controlling RNA virus infections in plants. CRISPR-mediated resistance against RNA viruses was first reported by Zhang T. et al. (2018) in transgenic Arabidopsis and tobacco plants. In the study, gRNA/FnCas9 cassettes were designed to target and attack various regions in the RNA genome of *Tobacco mosaic virus (TMV)* and *Cucumber mosaic virus*. Transgenic plants with gRNA/FnCas9 constructs were found to have significantly less viral accumulation (40 to 80%) relative to the control plants. Recently discovered RNA-targeting endonuclease, Cas13, has also been used to develop plant resistance against RNA virus infection (Aman et al., 2018). For instance, tobacco plants overexpressing gRNA/Cas13a successfully targeted and inhibited the replication of *Turnip mosaic virus* (Zhang et al., 2019). Similar approaches were efficiently used for the generation of resistance in potato against *Potato virus Y* (Zhan et al., 2019) and in rice for resistance to *Southern black-streak* and *stripe mosaic viruses* (Zhang et al., 2019).

Transgenic expression of the CRISPR constructs in transgenic plants and targeting the host susceptible factors is another strategy that was also used for viral resistance. Some translation initiation factors (eIF4E, eIF(iso)4E, and eIF4G) or their isoforms are required for replication and infection of RNA viruses. Therefore, the inactivation of these susceptibility factors in plants could be used to induce resistance to a virus without damage to the plant due to their functional redundancy between the different isoforms (Cao et al., 2021). CRISPR/Cas9 has been utilized to introduce mutations into these translation initiation factors in rice and tomato (Shimatani et al., 2017). Using the same editing technique, Bastet et al. (2019) introduced a single substitution mutation into eIF4E in the plant host genome. Both studies demonstrate that mutations of these susceptibility factors are sufficient to generate viral resistance in hosts against the potyviruses *Clover yellow vein virus*.

3.2 Genome editing for bacterial resistance

Genome editing in plants to develop resistance against bacterial diseases is still limited in application (Table 2). One of the main approaches taken to develop genome-edited plants resistant to

TABLE 1 Crop genome editing for viral resistance.

Crop	Target	Genetic Changes	Method	Status	Reference
C. sativus.	*Cucumber vein yellowing virus CVYV) Zucchini yellow mosaic virus (ZYMV) and Papaya ring spot mosaic virus-W (PRSMV)*	Base editing (A-G) of susceptibility factor (eIF 4E)	*Agrobacterium mediated gRNA/cas9 transfer into arabidobsis and selection of non-transgenic mutants in T3*	Broad virus resistance in non-transgenic cucumber	Chandrasekaran et al. (2016)
H. vulgare	*Wheat dwarf virus (WDV)*	Knockout the MP, RP, CP and IR of WDV	*Agrobacterium-mediated transient expression in barley*	Efficient viral resistance in barley	Kis et al. (2019)
M. esculenta	*Cassava brown streak virus (CBSV) and Ugandan cassava brown streak virus (UCBSV)*	Mutations in cassava susceptibility factor (eIF4E) of cassava	*Agrobacterium mediated stable gRNA/cas9 transfer and selection of mutants in cassava*	Suppressed disease symptoms and reduced virus titre in mutant cassava roots compared to WT	Gomez et al. (2019)
M. esculenta	*African Cassava Mosaic Virus (ACMV)*	Knockout in Rep and *MP* genes of ACMV	*Agrobacterium mediated stable expression of Cas9 protein together with gRNA*	Fails to confer effective resistance to ACMV in cassava and viral mutant escape	Mehta et al. (2019)
Musa spp.	*Banana Streak Virus (BSV)*	Targeting the ORF and IR in BSV	*Generation of transgenic banana with agrobacterium mediated gRNA/Cas9 transfer*	Full resistance in transgenic banana to endogenous BSV	Tripathi et al. (2019)
N. benthamiana	*Tomato yellow leaf curl virus (TYLCV)*	Knockout the viral IR, MP and REP coding region	*Stable overexpression of CAs9 in tobacco and transient expression of gRNA in tobacco*	Delayed and reduced accumulation of viral DNA, significantly attenuating symptoms of TYLCV infection in tobacco.	Ali et al. (2015b)
N. benthamiana	*Bean yellow dwarf virus (BeYDV)*	Knockout the viral LIR	*Transient expression assay with Agrobacterium*	Reduced virus load and disease symptoms in BeTDV treated tobacco	Baltes et al. (2015)
N. benthamiana	*Beet severe curly top virus (BSCTV)*	Knockout the coding and non-coding parts of BSCTV genome	*Stable and transient expression of gRNA/Cas9 in Arabidobsis and tobacco*	Strong reduction in viral load and disease symptoms in tobacco and *Arabidopsis*	Ji et al. (2015)
N. benthamiana and S. lycopersicum	*Tomato yellow leaf curl virus (TYLCV)*	Knockout the CP and Rep of TYLCV	*Agrobacterium-mediated stable gRNA/Cas9 transfer into tomato and tobacco*	Low accumulation of TYLCV in tomato and tobacco transgenic plants	Tashkandi et al. (2018)
N. benthamiana	*Cucumber mosaic virus (CMV) or tobacco mosaic virus (TMV)*	Targeting the RNA viruses	*Plants expressing FnCas9 and sgRNA specific for the RNA viruses*	Significantly attenuated virus infection symptoms and inheritable reduced viral accumulation in plants	Zhang T. et al. (2018)
N. benthamiana	*Chilli leaf curl virus (ChiLCV)*	Multiple targeting the genes of ChiLCV	*Agrobacterium transient assay in tobacco*	Resistant to ChiLCV with reduced viral accumulation	Roy et al. (2019)
N. benthamiana	*Cabbage leaf curl virus (CaLCuV)*	Knockout the viral IR and REP coding region	*Agrobacterium mediated transient expression of gRNA/Cas9 constructs*	Complete resistance to CuLCuV infection in transgenic tobacco	Yin et al. (2019)
O. sativa	*Rice tungro spherical virus (RTSV) and Rice tungro bacilliform virus (RTBV)*	Knockout in initiation factor 4 gamma gene (*eIF4G*)	*Agrobacterium-mediated transformation of gRNA/Cas9 into rice immature embryos*	In-frame mutations in one conferred resistance to RTSV and RTBV in rice	Macovei et al. (2018)
S. tuberosum L.	*Potato virus Y (PVY)*	Targeting conserved regions in expressed genes of PVY strains.	*Agro-infiltration of tobacco leaves and generation transgenic potato plants with LshCas13a/sgRNA*	Suppressed PVY accumulation and disease symptoms in transgenic potato	Zhan et al. (2019)
B.vulgaris	*Beet curly top Iran virus(BCTIV)*	Multiple targeting of the expressed genes of BCTIV	*Agrobacterium-mediated transient expression of gRNA/Cas9 in sugar beet leaves*	Full viral resistance in sugar beet	Yıldırım et al. (2023)

TABLE 2 Crop genome editing for bacterial resistance.

Crop	Target	Genetic Changes	Method	Status	Reference
C. maxima	*Xanthomonas citri subsp. citri (Xcc)*	EBE region of the LOB1 promoter in Pummelo	*Agrobacterium-mediated transformation of Pummelo epicotyls and obtaining T0*	Generation canker-resistant citrus varieties by mutation of the EBE	Jia and Wang (2020); Jia et al. (2017)
C. sinensis Osbeck	*Xanthomonas citri subsp. citri (Xcc)*	CRISPR/Cas9-targeted mutation in CsLOB1 promoter in citrus	*Agrobacterium mediated gRNA/Cas9 transfer and generation homozygous mutant citrus explants*	Promoter editing of CsLOB1 alone was sufficient to enhance citrus canker resistance in citrus.	Peng et al. (2017)
C. sinensis Osbeck	*Xanthomonas citri subsp. Citri (Xcc)*	Mutation and loss of function in *CsWRKY22*	*Agrobacterium mediated transformation of gRNA/Cas9 into epicotyl segments of orange*	Mutant orange plants showed decreased susceptibility to citrus canker	Wang et al. (2019)
M. balbisiana	*Xanthomonas campestris pv. musacearum (Xcm)*	Mutation in downy mildew resistance 6 (DMR6)	*gRNA/Cas9 was introduced into the embryogenic cell suspension through Agrobacterium-mediated transformation*	Musa dmr6 transgenic mutants of banana showed enhanced resistance to BXW, and did not show any detrimental effect on plant growth	Tripathi et al. (2021)
M. domestica	*Erwinia amylovora*	Mutation in apple *DIPM-1, DIPM-2* and *DIPM-4*	*Delivery of CRISPR/Cas9 ribonucleoproteins to the protoplast of apple cultivar*	Resistance to fire blast disease in non-transgenic but mutant apple lines	Malnoy et al. (2016)
O. sativa	*Xanthomonas oryzaepv. Oryzae (Xoo)*	Knockdown of the *Os8N3* in rice	*Stable transmission of CRISPR/Cas9-mediated Os8N3 gene editing without the transferred DNA)*	Transmission of mutations to generations, and enhanced resistance to Xoo in homozygous mutants.	Kim et al. (2019)
O. sativa	*Xanthomonas oryzae pv. oryzae (Xoo)*	Mutations in EBE of three promoters of *SWEET11, SWEET13* and *SWEET14*	*Promoter mutations were simultaneously introduced into the rice with Agrobacterium mediated transfer of gRNA/Cas9*	Stable transgenic rice lines indicated robust, broad-spectrum resistance to Xoo.	Oliva et al. (2019); Xu et al. (2019)
O. sativa	*Xanthomonas oryzae pv. oryzae (Xoo)*	Mutation in EBEs of *OsSWEET14* gene	*Biolistic technology was used to deliver gRNA/Cas9 into embryogenic calli of the rice*	Enhanced resistance locally isolated virulent Xoo strains	Zafar et al. (2020)
O. sativa	*Xanthomonas oryzaepv. Oryzae (Xoo)*	Mutation and loss of function in *OsSWEET14*	*Agrobacterium mediated stable expression of Cas9 protein together with gRNA*	Mutant rice confers strong resistance to African Xoo and Asian Xoo starins	Zeng et al. (2020)
S. lycopersicum	*P. syringae, P. capsici and Xanthomonas spp*	Loss of function mutation in *SlDMR6-1* gene	*Agrobacterium mediated transformation of gRNA/Cas9 tomato*	Mutants do not have detrimental effects on growth and had multiple disease resistance *P. syringae, P. capsic*i and *Xanthomonas* spp.	Paula de Toledo Thomazella et al. (2016)

pathogenic bacteria is by the knockout of susceptibility (*S*) gene/s (Zaidi et al., 2018). These genes are transcription factors that bind to a sequence-specific promoter region and are known as effector-binding elements (EBEs). A classic example of an *S* gene is the *Mildew Resistance Locus O* (MLO) which was first associated with powdery mildew (PW) susceptibility in barley decades ago (Jørgensen, 1992). An *S* gene related to bacterial infection, *SWEET* (Sugar Will Eventually Be Exported Transporter) gene in rice is related to susceptibility to *Xanthomonas oryzae* pv. oryzae (Xoo) (Antony et al., 2010). Furthermore, *Citrus sinensis lateral organ boundary 1* (CsLOB1) gene first identified as susceptibility gene for citrus bacterial canker that caused by *Xantomonas citri* subsp. citri (Xcc) recently found to play a regulatory role with activity in cell wall remodeling and in cytokinin and brassinosteroid hormone pathways. In favor of this statement, RNAi-mediated silencing of the CsLOB1 gene developed resistance to canker disease in various citrus species (Zou et al., 2021). Contrary to silencing, overexpressing of Gretchen Hagen3 (GH3.1 and GH3.1L) genes involved in auxin signaling in citrus significantly reduced susceptibility to *Xantomonas citri* subsp. citri (Zou et al., 2019). *S* gene GE approaches such as TALEN and CRISPR/Cas9 technologies were later used to mutate effector-binding sites within the SWEET promoter and develop resistant to Xoo in rice and tomato (Zafar et al., 2020; Zeng et al., 2020; Luo et al., 2021). Similarly, in a recent study, *DOWNY MILDEW RESISTANCE 6* (*DMR6*) was mutated using CRISPR/Cas9 mediated GE to successfully produce mutant banana and tomato plants resistant against *Xanthomonas campestris* (Xcm) and other pathogenic microbes (Wang et al., 2019; Tripathi et al., 2021). Luo et al. (2021) reported immunity of rice plants to bacterial blight Xoo by employing the CRISPR/Cas9 GE system to knockout *OsPrx30* a CIII Prx precursor.

3.3 Genome editing for fungal resistance

Loss of *S* gene function can provide more durable fungal resistance in other crop plants (Table 3). Functional knockouts of *StDND1*, *StCHL1*, and *StDMR6-1* susceptibility genes using CRISPR/Cas9 system generated potatoes with increased resistance against late blight (Kieu et al., 2021). Simultaneous modification of three homologues of *TaERD1* gene utilizing CRISPR/Cas9

TABLE 3 Crop genome editing for fungal resistance.

Crop	Target	Genetic Changes	Method	Status	Reference
C. lanatus	*Fusarium oxysporum.*	Loss-of-function in Phytosulfokine1 *(ClPSK1)* in watermelon	*Transformation of gRNA/Cas9 to watermelon through Agrobacterium tumefaciens-mediated transformation*	Loss-of-function rendered watermelon seedlings more resistant to infection by *F. oxysporum.*	Zhang M. et al. (2020)
C. papaya	*P. palmivora*	Mutation on Extracellular cystatin-like cysteine protease inhibitor (PpalEPIC8) of papaya	*PpalEPIC8 mutants were generated using CRISPR/Cas9-mediated gene editing via Agrobacterium-mediated transformation*	Reduced pathogenicity during infection	Gumtow et al. (2018)
G. hirsutum	*Verticillium dahliae*	Indel mutations in negative defence gene *(Gh14-3-3d)* of cotton	*Agrobacterium-mediated transformation of gRNA/Cas9 into cotton*	Higher and heritable resistance to *Verticillium dahliae* infestation in mutant cottons	Zhang Z. et al. (2018)
O. sativa	*M. oryzae*	Mutation in rice ERF Transcription Factor Gene *OsERF922*	*Agrobacterium-mediated transformation of the embryogenic calli of rice*	Enhanced resistance in mutant rice to *M. oryzae* in subsequent generations	Wang et al. (2016)
S. lycopersicum	*Powdery Mildew Resistance 4 (SlPMR4)*	Knock-out of the tomato *SlPMR4* gene	*Transferring CRISPR/Cas9 construct containing four single-guide RNAs (sgRNAs)to target SlPMR4*	Haustorial formation and hyphal growth were diminished but not completely inhibited in the mutants	Santillán Martínez et al. (2020)
S. lycopersicum	*Fusarium oxysporum f.* sp. *Lycopersici,*	Mutation in *SlymiR482e-3p*, a member of the *miR482/2118* superfamily in tomato, negatively regulating the resistance	*Agrobacterium transfer of gRNA/ Cas9 into susceptible tomato cultivar*	Enhanced resistance to tomato wilt disease in edited plants	Gao et al. (2021)
S. tuberosum	*Phytophthora infestans*	Tetra-allelic deletion of StDND1, StCHL1, and StDMR6-1 in potato	*Agrobacterium mediated transfer of multiple gRNA/Cas9 in to potato*	Editing confers increased late blight resistance in potato	Kieu et al. (2021)
T. cacao	*Phytophthora tropicalis*	Deletions in Non-Expressor of Pathogenesis-Related 3 *(TcNPR3)* gene, a suppressor of the defence response	*Agrobacterium was used to introduce a CRISPR/Cas9 system into leaf tissue*	The edited tissue exhibited an increased resistance to infection with the cacao pathogen *Phytophthora tropicalis*	Fister et al. (2018)
T. aestivum	*Powdery mildew*	Indel mutations at the wheat Mildew-resistance locus (MLO)	*Wheat protoplasts transformation with TALEN and CRISPR vectors*	TALEN and CRISPR-induced mutation at TaMLO homeologs, confers heritable broad-spectrum resistance to powdery mildew.	Wang et al., 2014
T. aestivum	*Blumeria graminis f.* sp. *tritici (Bgt)*	Simultaneous modification of the three homologs of wheat enhanced disease resistance1 *(TaEDR1)*	*Biolistic transformation of gRNA/ Cas9 plasmids into wheat immature embryos*	Mutant wheats were resistant to powdery mildew and did not show mildew-induced cell death.	Zhang Y. et al. (2017)
V. vinifera	*Erysiphe necator and Plasmopara viticola*	Editing the *DM* and *PM* susceptibility genes in different grapevine clones	*CRISPR/Cas9 technology was used to edit DM and PM susceptibility genes*	Multiple resistance against grape wine powdery mildew and downy mildew	Giacomelli et al., 2018
V. vinifera	*Oomycete pathogen Plasmopara viticola*	Loss-of-function mutations in *grapevine pathogenesis-related 4 (PR4)*	*Agrobacterium mediated gRNA/Cas delivery into Thompson Seedless*	The VvPR4b knockout lines had increased susceptibility and disease symptoms of downy mildew in mutant grapevine	Li M. Y. et al. (2020)

procedure enhanced powdery mildew resistance in wheat, caused by the biotrophic pathogen *Blumeria graminis* f. sp. *tritici* (Bgt) (Zhang Y. et al., 2017). TALENs system was used to modify Mildew resistant LOCUS (MLO) encoding proteins which repress powdery mildew defence in wheat. TALEN-induced mutation triggered heritable broad-spectrum resistance against powdery mildew disease (Wang et al., 2014). Transgene-free powdery mildew-resistant tomato variety was generated by deleting 48 bp region from *SlMLO1* locus utilizing CRISPR/Cas9 technology. The resulting plants were indistinguishable from naturally occurring mutations having the same phenotypic characteristics (Nekrasov et al., 2017). Another study was performed on tomato *Powdery Mildew Resistance 4* (*PMR4*) gene mutagenesis through CRISPR/Cas9 which resulted in mutants with reduced but not complete loss of susceptibility to powdery mildew pathogen *Oidium neolycopersici* (Santillán Martínez et al., 2020). To define the functions of SlymiR482e-3p gene in response to tomato wilt disease, caused by the *Fusarium oxysporum* f. sp. *lycopersici* fungus, CRISPR/Cas9 was used to knock-out the gene in a disease susceptible tomato cultivar. The resulting tomato mutants exhibited significant disease reduction (more than 90%) in SlymiR482e-3p levels and increased resistance to the necrotrophic pathogen displaying the

same phenotypic traits with the control plants (Gao et al., 2021). In *Gossypium hirsutum*, simultaneous editing of two *Gh14-3-3d* gene copies through CRISPR/Cas9 technology led to enhanced transgene-free resistance to *Verticillium dahliae* in allotetraploid cotton (Zhang Z. et al., 2018). *Agrobacterium*-mediated transient transformation was used to introduce CRISPR/Cas9 components into cacao leaves and cotyledon cells targeting *Non-Expressor of Pathogenesis-Related 3* (*TcNPR3*) gene, a suppressor of the defense response. The edited tissues exhibited enhanced immunity against *Phytophthora tropicalis* which is a widespread fungal pathogen (Fister et al., 2018). Developing *Fusarium oxysporum* (FON) resistant watermelon varieties by traditional breeding methods is hampered by the limited FON-resistant germplasm. Knockout of *Clpsk1* gene in watermelon through CRISPR/Cas9 system conferred resistance to FON, and thus established a base to develop disease-resistant germplasm in watermelon (Zhang M. et al., 2020).

As discussed previously plants have evolved complex defense mechanisms including plant hormones such as abscisic acid, salicylic acid, jasmonic acid and ethylene. Plant ethylene responsive factors (ERF) play roles in various biotic stress responses. The ethylene responsive factor *OsERF922* was edited using CRISPR/Cas9 which led to enhanced blast resistance in rice (Wang et al., 2016). In summary, gene editing technologies can offer robust and durable resistance against the most destructive fungal pathogens confronted in crop production worldwide.

3.4 Genome editing for resistance of crops to parasitic weeds

Plants are autotrophic organisms using light as energy for converting carbon into carbohydrates by photosynthesis. On the other hand, some other plants have evolved specialized organs (haustorium) which attach forming vascular connections with autotrophic plants in order to absorb their water and nutrients. This heterotrophic lifestyle is known as parasitic plants/weeds and has a profound negative impact on many important crops and trees affecting these ecological systems (Hu et al., 2020). The existence of parasitic plants in lower diversified agrological systems can cause yield losses and make some land uncultivable (Fernández-Aparicio et al., 2020). Weeds tend to compete with crops for water, nutrients, and light sources. However, parasitic weeds' haustorial connections to either the xylem or phloem directly extract water and nutrients from host plants and cause permanent damage to the crops' life cycle (Albert et al., 2020). Traditional weed management methods tend to be ineffective, expensive and labor-intensive. Parasitic weeds generally produce large numbers of small seeds that make it difficult to detect and eradicate contamination of the soil or the crop seeds before parasitism are established. The seeds of parasitic plants have long dormancy and viability in soils and germinate after receiving the host signals (Delavault, 2020).

The discovery of the terpenoid lactones in crops (e.g., strigolactones (SLs) and sesquiterpene lactones (STLs), Xie et al., 2010; Chadwick et al., 2013) is a milestone in understanding the responses of parasitic weeds and to their hosts. Host roots synthesize trace amounts of secondary metabolites which have several important physiological processes in host plants from shoot branching to arbuscular mycorrhizal symbiosis. Terpenoid lactones were then realized to be the germination stimulants for several obligate parasitic species, including broomrapes (e.g., *Orobanche* and *Phelipanche* spp.) (Raupp and Spring, 2013; Cheng et al., 2017). The seeds of these parasitic plants germinate when they receive terpenoid lactone signals from their hosts. Thus interactions between parasitic weeds and hosts have evolved in a very specific way dependent on the detection of the presence of STLs or SLs by parasitic weeds and coordinate their germination and development with the host's lifecycle (Spring, 2021). Reducing the quantity of such stimulant exuded by host plants was always considered to be a key factor for the host resistance achieved by inhibition of parasitic weed seed germination. CRISPR, and RNAi mediated gene silencing strategies have been used to block strigolactones (SLs) synthesis in hosts (Vogel et al., 2010; Kohlen et al., 2012; Aly et al., 2014; Dubey et al., 2017; Butt et al., 2018; Bari et al., 2019; Wakabayashi et al., 2019). In this way, the germination of seeds of parasitic plants was suppressed and almost complete resistance to parasitic weeds was achieved in genome-edited host plants.

3.5 Recent advances in genome editing and new potential applications for plant pathogen resistance

Base editors enable single-nucleotide changes in the genomes without cutting or removing the nucleic acid backbone. CRISPR-Mediated base editing (CBE) used a single-stranded DNA-specific cytidine deaminase fused to an inactivated Cas9 (dCas9) to convert a cytosine (C)-guanine (G) base pair to thymine (T)-adenine (A) in the target region with the help of sgRNA (Li et al., 2023). CBE have lots of potential and theoretical application that can be used for disease resistance in plants. For instance, a CBE can convert C to T (G to A in the opposite strand) precisely, turning glutamine (CAA and CAG), arginine (CGA), and tryptophan (TGG) codons into stop codons (Kuscu et al., 2017). If this precise substitution (generating a nonsense mutation) occurs in the gene of interest, it will cause premature termination of translation and abort the gene's function. It can also be used to alter splicing mechanisms in plant species. The splicing of intronic regions highly depends on conserved 5′GT and 3′AG sequences. Theoretically, if the conserved sites mutate, it will interfere with mRNA splicing, cause mRNA mis-splicing, and eventually disrupt gene function. In addition to this loss-off function application; CBE can be also used for the gain of function in plants. In this technique introducing a targeted point mutation in a gene turns a nonfunctional SNP into a functional one. The best example of CBE-based gain of function is *Acetolactate synthase* (ALS) gene which is a key enzyme in the biosynthesis of branched-chain amino acids, making it an effective target for developing herbicides. CBE was exploited to target wheat mutants ALS with a change of C-to-T conversion at the conserved Pro174 residue and they showed herbicide resistance (Zong et al., 2017). Similarly, CBE was used to create a series of missense mutations in the *OsALS* to confer herbicide tolerance in rice

(Zhang R. et al., 2020). ALS has also been successfully edited in other species (Chen et al., 2017; Tian et al., 2018; Veillet et al., 2019; Jiang et al., 2020). These base editor systems have been also effectively used to enhance plants' resistance to pathogens. For instance, previous studies have indicated that a single amino acid substitution at position 441 of the recessive allele of the Pi-d2 gene resulted in the loss of resistance to rice blast (*M. oryzae*). CBE technology was successfully used to introduce a G-to-A substitution in a recessive allele of Pi-d2. The deduced protein contained an amino acid substitution, which recovered the resistance of rice to *Magnaporthe oryzae* (Ren et al., 2018). In another study, Wang et al. (2020) used the same system to target the effector binding element within the promoter of the OsSWEET14 gene in rice. The base-edited mutant rice exhibited high resistance to the leaf blight fungus. Using the CBE technique, Bastet et al. (2019) introduced a single substitution mutation into eIF4E in the plant host genome and mutations of this susceptibility factor were found to be sufficient for resistance to *potyviruses clover yellow vein virus*.

In recent years, enzymatically inactive mutant of Cas9 (dead or deactivated Cas9 -dCas9) was developed in which its endonuclease activity is non-functional. The applications of CRISPR/dCas9 have expanded and diversified in recent years (Moradpour and Abdulah, 2020). Originally, dCas9 was used as a CRISPR/Cas9 re-engineering tool that enables targeted expression of any gene or multiple genes through recruitment of transcriptional effector domains (promoters) without introducing irreversible DNA-damaging mutations (Figure 2). dCas9 started to become a powerful tool for targeted inhibition of gene transcription in plants. dCas9 can easily directed with sgRNAs to the promoter regions of the genes and functions as a repressor or block for the transcriptional machinery, a phenomenon called CRISPR interference (CRISPRi). CRISPRi has been reported to be used for effective, stable RNA-guided transcriptional suppression of a target gene in several plant species (Larson et al., 2013; Qi et al., 2013). sgRNA-guided CRISPR activation (CRISPRa or CRISPR-Act) systems have also been developed in plants for increased expression of target genes. In this system, various gene activator proteins were fused to the dCas9 and directed to the promoter region of the target genes with sgRNAs. Binding of CRISPRa to the target promoter region up-regulated expression of the gene of interest in plants (Tiwari et al., 2012; Piatek et al., 2015; Li et al., 2017; Lowder et al., 2017a; Lowder et al., 2017b) (Figure 2). Both CRISPRa and CRISPRi technologies have not been utilized for the improvement of plant resistance to pathogens yet. However, they have a large potential and flexibility that can be used for R gene-mediated resistance in plants instead of S gene-dependent loss of function approaches.

Genome editing is widely applied via stable integration of gRNA/Cas9 construct with selective marker gene to plants' genome. However, transgene integration in plant genomes raises important legislative concerns regarding genetically modified plants. In order to obtain transgene-free edited plants, it is necessary for the integrated foreign DNA to segregate out via selfing or crossing with wild-type plants (Gao et al., 2021). This is a labor intensive and time-consuming process, and thus not suitable for several plant species. Genome editing by using CRISPR ribonucleoproteins (RNPs) has become an attractive approach for many crop species with many advantages. In this system, a ribonucleoprotein (RNP) complex consisting of Cas9 protein and single guide RNA (sgRNA) directly delivered to protoplast cell culture via bombardment, polyethylene glycol-mediated transfection or electro-transfection. RNP-mediated genome editing can be achieved shortly after cell transfection because transcription or translation is not required. RNP complex is degraded in the cell and transgene free mutant plant lines could be obtained after regeneration. This system would become a powerful and widespread method for genome editing due to its advantages of DNA/transgene-free editing, minimal off-target effects, and reduced toxicity due to the rapid degradation of RNPs and the ability to titrate their dosage while maintaining high editing efficiency. Although RNP-mediated genetic engineering has been demonstrated in many plant species, its editing efficiency remains modest, and its application in many species is limited by difficulties in plant regeneration and selection. Although RNP-mediated genetic engineering has been demonstrated in many plant species (Zhang Y. et al., 2021), its editing efficiency in terms of pathogen resistance in plants remain to be tested.

4 Impacts – risks, challenges and future perspectives

Genome editing is offering new tools and opportunities for the improvement of plant disease resistance. The development of efficient methods for its wider application in resistance breeding has the potential to create a significant impact on crop cultivation in the future. However, as with all things new, it will face some challenges to be overcome, and create potential risks that have to be taken into account.

4.1 Impacts on crop improvement – advantages and limitations

Plant breeding for the production of new resistant varieties using classical approaches has had limited success (Ahmad et al., 2019) due to the potential of pathogens, through recombination and/or mutation and the development of novel genotypes that are no longer sensitive to resistance genes. G E enables the production of desired pathogen/pest resistance in plants that could supplement traditional or molecular breeding methods and reduce breeding cycles. So far, the successful application of genome editing in pathogen/pest resilience and its introduction into crops has been limited by the lack of information on genome sequences in crop plants and the characterization of potential target genes. Fortunately, numerous species have been fully sequenced in last decades, enabling genome editing of many crops. Crop genetic studies have described details of crop immunity, and now identified larger numbers of potential targets for control of pathogens.

A range of plant defence mechanisms can be used by breeders to protect plants. The R gene-type of disease resistance has been exploited in traditional plant breeding and generally, it is preferred over immunity systems based on PTI as it is a qualitative resistance

easier to select. However, it is less durable and pathogen populations easily adapt to overcome disease resistance (Collinge, 2020; Li M. Y. et al., 2020; Li W. et al., 2020). Pathogen resistance obtained through *R*-genes is limited in use as *R*-gene conferred resistance is generally pathogen race specific and is overcome by the evolution of new races. Hence, susceptibility regulators of disease resistance, (*S*-genes), provide better targets for GE (Yin and Qiu, 2019). They have emerged as an alternative to *R*-genes, as *S*-genes conferring resistance are recessively inherited and editing in 'S' alleles through CRISPR/Cas9 exhibits more broad-spectrum and durable forms of resistance than resistance genes (R-genes) against pathogens (Yildirim et al., 2012; Van Schie and Takken, 2014) and provide crop resistance that has the potential to be more persistent in the field. Using GE techniques such as CRISPR-Cas9 or RNAi can remove or inactivate these genes and impair the pathogens' ability to cause disease (Lapin and Van den Ackerveken, 2013). S-gene mutants can be produced in most crops without considering species barriers due to the functional conservation of S-genes across crop species. Further advances in molecular studies of main crops will also enable the discovery of novel S-genes, thus providing additional targets for GE. However, S-genes are also involved in other plant physiological processes so their inactivation could disrupt crop development. This factor may hamper the application of S-gene editing in crop improvement (Yin and Qiu, 2019).

Progress in understanding the specific processes involved in pathogen-host interactions is expected to pave the way to the employment of gene drives for the creation of crops that are immune to certain pathogens and pests, and no longer support pathogen growth (Hefferon and Herring, 2017). Gene drive systems applied for eradicating malaria vector mosquitoes (Kyrou et al., 2018) could be used as a model for control of sexually inheriting crop pests and pathogens. However, the application of gene drives could also lead to changes in entire pest and pathogen communities thus affecting current ecosystems which need to be considered before wider application of this technology in crop improvement (Hefferon and Herring, 2017). Finally, the recent publication and subsequent retraction of the article of Zhang T. et al. (2021) and Zhang et al. (2022) reporting the design of a gene drive based on CRISPR Cas9 that targets specific genetic elements in *Arabidopsis*, shows that the application of gene drives, although promising, is still not in ready for wide use in crop improvement.

4.2 Risks and challenges

If gene editing involves transgenics (*i.e.*, SDN-3) then there is a consensus that the products are considered as GMOs by most regulatory authorities or as Novel Plants by the Canadian authorities. SDN-1 and SDN-2 plants usually do not contain foreign DNA and so are not regulated as GMOs in many/most countries, an exception being the European Union (EU) (Rostocks, 2021). In many countries non-transgenic gene editing is considered a development of conventional breeding and so regulations are being developed on this basis (Jenkins et al., 2021). However, the European Commission is now proposing that plants that could be created using conventional breeding techniques are exempt from the EU GMO regulations (European Commission, 2023). This would include many genome edited plants from SDN-1 and SDN-2 as well as some cisgenic types. Thus there is some convergence of the regulations on GE plants.

As in all plant breeding processes, unintended and off target effects can also occur in gene editing, though it is argued that gene editing has higher levels of precision and targeting so that they will occur much less frequently than in conventional and other types of mutation breeding, as well as the transformation techniques using DNA as the transfecting agent (Okita and Delseny, 2023). In addition, there may be reduced genetic stability in GE plants in target or associated *loci*. These effects may compromise the efficacy and durability of enhanced pest and disease resistance. It is therefore important that plant breeders test for and identify any pleiotropic, off target or stability effects using both molecular techniques and phenotypic, field studies.

Some attempts to develop viral resistance using CRISPR have not been successful and presented some of the disadvantages of CRISPR. Mehta et al. (2019) was not able to induce resistance to *African cassava mosaic virus* in GM cassava plants that overexpressed gRNA/Cas9 constructs targeting the viral transcription activator and replication enhancer protein. Similar results were also recorded when the CRISPR/Cas9 system was used to block coding sequences of *TYLCV, MeMV*, and *Cotton leaf curl Kokhran virus* (Ali et al., 2016). Instead, both studies found that CRISPR editing produced new mutant variants which were probably due to repair post-cleavage. Thus, GE may present some risks, due to the production of new virus variants.

Procedures for risk assessment of GE plants have been proposed by Eckerstorfer et al. (2021) and Lema (2021) and discussed by EFSA (Naegeli et al., 2020). Obviously, changes to nutritional quality of plants should be assessed but Eckerstorfer and co-workers (Eckerstorfer et al., 2021) also stressed that novel or enhanced traits of GE plants should be considered for their environmental impacts. Factors to be considered include any non-target effects, including considering changes in pathogenicity and weediness of pathogens associated with GE crops and consequences of changes to fitness and invasiveness of GE plants and hybridizing species.

4.3 Conclusions and future challenges

Due to its various advantages, CRISPR/Cas technology has become the technology of choice in wide aspects of scientific research for a short time period. However, there are still bottlenecks and challenges for its wider implementation and usage. In most life-science laboratories in the world, this technology has found its place in fundamental research, prevalently in animal cells in comparison to plants. The future research directions of GE in plants should evolve in aspects of gene delivery, resolving modalities to expand high throughput editing strategies, discovering new Cas enzymes to lower the limitations in specific gene targeting and enhancing the regenerative capacity and stability of transformation effects. One of the future research directions in genome editing, that will be of crucial importance in increasing current low transformation efficacy is finding a solution

of breaking the recalcitrance in diverse plant species in tissue culture. To overcome this, alternative methods like virus-induced GE (VIGE) and nanotechnology-based GE, have recently been developed to avoid the need for *de novo* regeneration from tissue culture. However, to increase the adoption of these technologies, it will be important to overcome the limitations set by the size of the Cas enzyme (Cardi et al., 2023).

GE plants with improved pest and disease resistance have the potential to introduce more durable resistance and thus contribute towards more sustainable pest and disease management. More durable resistance mechanisms can be produced by down regulating S-genes and up regulating genes that identify pathogens, inhibit infection and reduce the virulence of pathogens by inhibiting development using genetic modifications, gene editing, RNAi and gene drives. Combining these mechanisms and managing levels of exposure to pests and pathogens in IPM can make major contributions to improving the sustainability of agricultural production, particularly in response to climate change, and to achieving Unite Nations Sustainable Development Goals and National/EU policy objectives for agriculture and the environment. Strategies for exploiting GE crops have been extensively reviewed by Bartlett et al. (2023). Of particular importance is the improvement of traits such as tolerance to biotic stresses and herbicides, self-compatibility to allow for self-pollination and inbreeding, lower content of toxic compounds as steroidal glycoalkaloids, browning free fruits and tubers, and improvements in starch quality (Tuncel and Qi, 2022).

Public perceptions and attitudes to the use of GE technologies for producing crops and foods are critical for the introduction of GE produce into food production and supply chains and require clear communication of the benefits and risks. These issues have been extensively discussed by several authors e.g., Strobbe et al. (2023) and Will et al. (2023).

It is important that appropriate and science-based policies and regulations are in place that allow rapid assessment of the risks of the products from these new breeding techniques. This will create preconditions for responsible usage of GE technology and its wider application. In addition, plant breeders and crop variety evaluators should be able assess the net contribution that new varieties can make to sustainable farming systems considering present and future requirements in relation to climate change and other externalities influencing food production and supply chains.

Author contributions

EA, JS, and DM developed manuscript outline. All authors contributed to writing and editing. KY prepared Figure and Tables. DM, EA, JS, and KY did final editing and manuscript preparation. All authors contributed to the article and approved the submitted version.

Funding

DM acknowledges funding by the Science, Technological Development, and Innovations of the Republic of Serbia, grant number 451-03-47/2023-01/200032, by the Science Fund of the Republic of Serbia through IDEAS project "Creating climate smart sunflower for future challenges" (SMARTSUN), grant number 7732457, by the European Commission through Twinning Western Balkans project CROPINNO, grant number 101059784. VG and MZ acknowledge funding from the Ministry of Science, Technological Development, and Innovation of the Republic of Serbia (contract no. 451-03-47/2023-01/200197). EA acknowledges funding from GREEN-IT Bioresources for Sustainability, ITQB NOVA, Av. da República, 2780-157 Oeiras, Portugal. This article is based upon work from COST Action PlantEd (CA18111), supported by COST (European Cooperation in Science and Technology). www.cost.eu.

Acknowledgments

DM acknowledges the support of Center of Excellence for Innovations in Breeding of Climate-Resilient Crops - Climate Crops, Institute of Field and Vegetable Crops, Novi Sad, Serbia.

Conflict of interest

Author JS is an independent agent operating as Sweet Environmental Consultants.

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

References

Aerts, N., Pereira Mendes, M., and van Wees, S. C. M. (2021). Multiple levels of crosstalk in hormone networks regulating plant defense. *Plant J.* 105, 489–504. doi: 10.1111/tpj.15124

Ahmad, S., Cheema, H. M. N., Khan, A. A., Khan, R. S. A., and Ahmad, J. N. (2019). Resistance status of Helicoverpa armigera against Bt cotton in Pakistan. *Transgenic Res.* 28, 199–212. doi: 10.1007/s11248-019-00114-9

Akın, M., Eyduran, S. P., Rakszegi, M., Yıldırım, K., and Rocha, J. M. (2023). "Statistical modeling applications to mitigate the effects of climate change on quality traits of cereals: A bibliometric approach," in *Developing Sustainable and Health Promoting Cereals and Pseudocereals* (Cambridge, Massachusetts, United States: Academic Press), 381–396.

Aksoy, E., Yildirim, K., Kavas, M., Kayihan, C., Yerlikaya, B. A., Calik, I., et al. (2022). General guidelines for CRISPR/Cas-based genome editing in plants. *Mol. Biol. Rep.* 49 (12), 12151–12164. doi: 10.1007/s11033-022-07773-8

Albert, M., Axtell, M. J., and Timko, M. P. (2020). Mechanisms of resistance and virulence in parasitic plant–host interactions. *Plant Physiol.* 185, 1282–1291. doi: 10.1093/plphys/kiaa064

Ali, Z., Abulfaraj, A., Idris, A., Ali, S., Tashkandi, M., and Mahfouz, M. M. (2015a). CRISPR/Cas9-mediated viral interference in plants. *Genome Biol.* 16, 1–1. doi: 10.1186/s13059-015-0799-6

Ali, Z., Abul-Faraj, A., Li, L., Ghosh, N., Piatek, M., Mahjoub, A., et al. (2015b). Efficient virus-mediated genome editing in plants using the CRISPR/Cas9 system. *Mol. Plant* 8, 1288–1291. doi: 10.1016/j.molp.2015.02.011

Ali, Z., and Mahfouz, M. M. (2021). CRISPR/Cas systems versus plant viruses: engineering plant immunity and beyond. *Plant Physiol.* 186 (4), 1770–1785. doi: 10.1093/plphys/kiab220

Ali, Z., Ali, S., Tashkandi, M., Zaidi, S. S. E. A., and Mahfouz, M. M. (2016). CRISPR/Cas9-mediated immunity to geminiviruses: differential interference and evasion. *Sci. Rep.* 6, 26912. doi: 10.1038/srep26912

Ali, Q., Yu, C., Hussain, A., Ali, M., Ahmar, S., Sohail, M. A., et al. (2022). Genome engineering technology for durable disease resistance: recent progress and future outlooks for sustainable agriculture. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.860281

Aly, R., Dubey, N. K., Yahyaa, M., Abu-Nassar, J., and Ibdah, M. (2014). Gene silencing of CCD7 and CCD8 in Phelipanche aEgyptiaca by tobacco rattle virus system retarded the parasite development on the host. *Plant Signal Behav.* 9 (8), e29376. doi: 10.4161/psb.29376

Aman, R., Ali, Z., Butt, H., Mahas, A., Aljedaani, F., Khan, M. Z., et al. (2018). RNA virus interference *via* CRISPR/Cas13a system in plants. *Genome Biol.* 19, 1–9. doi: 10.1186/s13059-017-1381-1

Amari, K., Huang, C., and Heinlein, M. (2021). Potential impact of global warming on virus propagation in infected plants and agricultural productivity. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.649768

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

Andolfo, G., and Ercolano, M. R. (2015). Plant innate immunity multicomponent model. *Front. Plant Sci.* 6. doi: 10.3389/fpls.2015.00987

Antony, G., Zhou, J., Huang, S., Li, T., Liu, B., White, F., et al. (2010). Rice xa13 recessive resistance to bacterial blight is defeated by induction of the disease susceptibility gene Os-11N3. *Plant Cell.* 22, 3864–3876. doi: 10.1105/tpc.110.078964

Backer, R., Naidoo, S., and van den Berg, N. (2019). The NONEXPRESSOR OF PATHOGENESIS-RELATED GENES 1 (NPR1) and related family: Mechanistic insights in plant disease resistance. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.00102

Balint-Kurti, P. (2019). The plant hypersensitive response: concepts, control and consequences. *Mol. Plant Pathol.* 20, 1163–1178. doi: 10.1111/mpp.12821

Baltes, N. J., Hummel, A. W., Konecna, E., Cegan, R., Bruns, A. N., Bisaro, D. M., et al. (2015). Conferring resistance to geminiviruses with the CRISPR-Cas prokaryotic immune system. *Nat. Plants* 1, 4–7. doi: 10.1038/NPLANTS.2015.145

Bari, V. K., Nassar, J. A., Kheredin, S. M., Gal-On, A., Ron, M., Britt, A., et al. (2019). CRISPR/Cas9-mediated mutagenesis of CAROTENOID CLEAVAGE DIOXYGENASE 8 in tomato provides resistance against the parasitic weed Phelipanche aEgyptiaca. *Sci. Rep.* 9, e11438. doi: 10.1038/s41598-019-47893-z

Bartlett, M. E., Moyers, B. T., Man, J., Subramaniam, B., and Makunga, N. P. (2023). The power and perils of *de novo* domestication using genome editing. *Ann. Rev. Plant Biol.* 74, 727–750. doi: 10.1146/annurev-arplant-053122-030653

Bastet, A., Zafirov, D., Giovinazzo, N., Guyon-Debast, A., Nogué, F., Robaglia, C., et al. (2019). Mimicking natural polymorphism in eIF4E by CRISPR-Cas9 base editing is associated with resistance to potyviruses. *Plant Biotechnol. J.* 17, 1736–1750. doi: 10.1111/pbi.13096

Berens, M. L., Berry, H. M., Mine, A., Argueso, C. T., and Tsuda, K. (2017). Evolution of hormone signaling networks in plant defense. *Annu. Rev. Phytopathol.* 55, 401–425. doi: 10.1146/annurev-phyto-080516-035544

Bhoi, A., Yadu, B., Chandra, J., and Keshavkant, S. (2022). Mutagenesis: A coherent technique to develop biotic stress resistant plants. *Plant Stress.* 3, 100053. doi: 10.1016/j.stress.2021.100053

Bisht, D. S., Bhatia, V., and Bhattacharya, R. (2019). Improving plant-resistance to insect-pests and pathogens: The new opportunities through targeted genome editing. *Semin. Cell Dev. Biol.* 96, 65–76. doi: 10.1016/j.semcdb.2019.04.00

Boschi, F., Schvartzman, C., Murchio, S., Ferreira, V., Siri, M. I., Galván, G. A., et al. (2017). Enhanced bacterial wilt resistance in potato through expression of arabidopsis efr and introgression of quantitative resistance from solanum commersonii. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.01642

Boutrot, F., and Zipfel, C. (2017). Function, discovery, and exploitation of plant pattern recognition receptors for broad-spectrum disease resistance. *Annu. Rev. Phytopathol.* 55, 257–286. doi: 10.1146/annurev-phyto-080614-120106

Butt, H., Jamil, M., Wang, J. Y., Al-Babili, S., and Mahfouz, M. (2018). Engineering plant architecture *via* CRISPR/Cas9-mediated alteration of strigolactone biosynthesis. *BMC Plant Biol.* 18, 174. doi: 10.1186/s12870-018-1387-1

Caarls, L., Pieterse, C. M. J., and van Wees, S. C. M. (2015). How salicylic acid takes transcriptional control over jasmonic acid signaling. *Front. Plant Sci.* 6. doi: 10.3389/fpls.2015.00170

Cao, Y., Zhou, H., Zhou, X., and Li, F. (2021). Conferring resistance to plant RNA viruses with the CRISPR/CasRx system. *Virol. Sinica* 1, 1–4. doi: 10.1007/s12250-020-00338-8

Carbonell, A., and Carrington, J. C. (2015). Antiviral roles of plant ARGONAUTES. *Curr. Opin. Plant Biol.* 27, 111–117. doi: 10.1016/j.pbi.2015.06.013

Cardi, T., Murovec, J., Bakhsh, A., Boniecka, J., Bruegmann, T., Bull, S. E., et al. (2023). CRISPR/Cas-mediated plant genome editing: outstanding challenges a decade after implementation. *Trends Plant Sci.* 28 (10), 1144–1165. doi: 10.1016/j.tplants.2023.05.012

Chadwick, M., Trewin, H., Gawthrop, F., and Wagstaff, C. (2013). Sesquiterpenoids lactones: benefits to plants and people. *Int. J. Mol. Sci.* 14 (6), 12780–12805. doi: 10.3390/ijms140612780

Chandrasekaran, J., Brumin, M., Wolf, D., Leibman, D., Klap, C., Pearlsman, M., et al. (2016). Development of broad virus resistance in non-transgenic cucumber using CRISPR/Cas9 technology. *Mol. Plant Pathol.* 17 (7), 1140–1153. doi: 10.1111/mpp.12375

Chen, Y., Wang, Z., Ni, H., Xu, Y., Chen, Q., and Jiang, L. (2017). CRISPR/Cas9-mediated base-editing system efficiently generates gain-of-function mutations in Arabidopsis Sci. *China Life Sci.* 60, 520–523. doi: 10.1007/s11427-017-9021-5

Cheng, X., Floková, K., Bouwmeester, H., and Ruyter-Spira, C. (2017). The role of endogenous strigolactones and their interaction with ABA during the infection process of the parasitic weed *Phelipanche ramosa* in tomato plants. *Front. Plant Sci.* 8, 392. doi: 10.3389/fpls.2017.00392

Collinge, D. B. (2020). Race specificity and plant immunity 12. *Plant Pathol. Plant Dis.* 12, 216. doi: 10.1079/9781789243185.0216

Collinge, D. B., Jensen, D. F., Rabiey, M., Sarrocco, S., Shaw, M. W., and Shaw, R. H. (2022). Biological control of plant diseases—what has been achieved and what is the direction? *Plant Pathol.* 71, 1024–1047. doi: 10.1111/ppa.13555

Collinge, D. B., and Sarrocco, S. (2022). Transgenic approaches for plant disease control: Status and prospects 2021. *Plant Pathol.* 71 (1), 207–225. doi: 10.1111/ppa.13443

Delavault, P. (2020). Are root parasitic plants like any other plant pathogens? *New Phytol.* 226, 641–643. doi: 10.1111/nph.16504

De Ronde, D., Butterbach, P., and Kormelink, R. (2014). Dominant resistance against plant viruses. *Front. Plant Sci.* 5. doi: 10.3389/fpls.2014.00307

Ding, S. W., and Voinnet, O. (2007). Antiviral immunity directed by small RNAs. *Cell* 130, 413–426. doi: 10.1016/j.cell.2007.07.039

Dodds, P. N., and Rathjen, J. P. (2010). Plant immunity: Towards an integrated view of plant pathogen interactions. *Nat. Rev. Genet.* 11, 539–548. doi: 10.1038/nrg2812

Doehlemann, G., Wahl, R., Horst, R. J., Voll, L. M., Usadel, B., Poree, F., et al. (2008). Reprogramming a maize plant: Transcriptional and metabolic changes induced by the fungal biotroph Ustilago maydis. *Plant J.* 56, 181–195. doi: 10.1111/j.1365-313X.2008.03590.x

Dubey, N. K., Eizenberg, H., Leibman, D., Wolf, D., Edelstein, M., Abu-Nassar, J., et al. (2017). Enhanced host-parasite resistance based on down-regulation of Phelipanche aEgyptiaca target genes is likely by mobile small RNA. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.01574

Eckerstorfer, M. F., Grabowski, M., Lener, M., Engelhard, M., Simon, S., Dolezel, M., et al. (2021). Biosafety of genome editing applications in plant breeding: Considerations for a focused case-specific risk assessment in the EU. *BioTech* 10, 1–14. doi: 10.3390/biotech10030010

Escobar-Bravo, R., Klinkhamer, P. G. L., and Leiss, K. A. (2017). Induction of jasmonic acid-associated defenses by thrips alters host suitability for conspecifics and correlates with increased trichome densities in tomato. *Plant Cell Physiol.* 58, 622–634. doi: 10.1093/pcp/pcx014

European Commission (2023). *Proposal for a Regulation of the European Parliament and of the council on plants obtained by certain new genomic techniques and their food and feed, and amending Regulation (EU) 2017/625* (Brussels). 5.7.2023 COM(2023) 411 final 2023/0226 (COD).

Fernández-Aparicio, M., Delavault, P., and Timko, M. P. (2020). Management of infection by parasitic weeds: A review. *Plants* 9, 1–26. doi: 10.3390/plants9091184

Fister, A. S., Landherr, L., Maximova, S. N., and Guiltinan, M. J. (2018). Transient expression of CRISPR/Cas9 machinery targeting TcNPR3 enhances defense response in theobroma cacao. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.00268

Franic, I., Prospero, S., Adamson, K., Allan, E., Attorre, F., Auger-Rozenberg, M. A., et al. (2022). Worldwide diversity of endophytic fungi and insects associated with dormant tree twigs. *Sci. Data* 9 (1), 8. doi: 10.1038/s41597-022-01162-3

Gao, Y., Li, S. J., Zhang, S. W., Feng, T., Zhang, Z. Y., Luo, S. J., et al. (2021). SlymiR482e-3p mediates tomato wilt disease by modulating ethylene response pathway. *Plant Biotechnol. J.* 19, 17–19. doi: 10.1111/pbi.13439

Giacomelli, L., Zeilmaker, T., Malnoy, M., Rouppe van der Voort, J., and Moser, C. (2018). "Generation of mildew-resistant grapevine clones via genome editing," in *XII International Conference on Grapevine Breeding and Genetics*, Vol. 1248. 195–200.

Gomez, M. A., Lin, Z. D., Moll, T., Chauhan, R. D., Hayden, L., Renninger, K., et al. (2019). Simultaneous CRISPR/Cas9-mediated editing of cassava eIF 4E isoforms nCBP-1 and nCBP-2 reduces cassava brown streak disease symptom severity and incidence. *Plant Biotechnol. J.* 17 (2), 421–434. doi: 10.1111/pbi.12987

Gong, P., Riemann, M., Dong, D., Stoeffler, N., Gross, B., Markel, A., et al. (2019). Two grapevine metacaspase genes mediate ETI-like cell death in grapevine defence against infection of Plasmopara viticola. *Protoplasma* 256, 951–969. doi: 10.1007/s00709-019-01353-7

Gosavi, G., Yan, F., Ren, B., Kuang, Y., Yan, D., Zhou, X., et al. (2020). Applications of CRISPR technology in studying plant-pathogen interactions: overview and perspective. *Phytopathol. Res.* 2, 21. doi: 10.1186/s42483-020-00060-z

Gumtow, R., Wu, D., Uchida, J., and Tian, M. A. (2018). *Phytophthora palmivora* extracellular cystatin-like protease inhibitor targets papain to contribute to virulence on papaya. *Mol. Plant-Microbe Interactions* 31 (3), 363–373. doi: 10.1094/MPMI-06-17-0131-FI

Gvozdenac, S., Dedić, B., Mikić, S., Ovuka, J., and Miladinović, D. (2022). "Impact of climate change on integrated pest management strategies" in *Climate change and agriculture*. Ed, N. Benkeblia (New Jersey, United States: John Wiley & Sons Ltd.), 311–372. doi: 10.1002/9781119789789.ch14

Hammond-Kosack, K. E., and Jones, J. D. G. (1997). Plant disease resistance genes. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 48, 575–607. doi: 10.1146/annurev.arplant.48.1.575

Hefferon, K. L., and Herring, R. J. (2017). The end of the GMO? Genome editing, gene drives and new frontiers of plant technology. *Rev. Agrarian Stud.* 7, 2369–2020-2023.

Hou, S., Liu, Z., Shen, H., and Wu, D. (2019). Damage-associated molecular pattern-triggered immunity in plants. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.00646

Hu, L., Wang, J., Yang, C., Islam, F., Bouwmeester, H. J., Muños, S., et al. (2020). The effect of virulence and resistance mechanisms on the interactions between parasitic plants and their hosts. *Int. J. Mol. Sci.* 21 (23), 9013. doi: 10.3390/ijms21239013

Hui, S., Shi, Y., Tian, J., Wang, L., Li, Y., Wang, S., et al. (2019). TALE-carrying bacterial pathogens trap host nuclear import receptors for facilitation of infection of rice. *Mol. Plant Pathol.* 20, 519–532. doi: 10.1111/mpp.12772

Hussain, B. (2015). Modernization in plant breeding approaches for improving biotic stress resistance in crop plants. *Turkish J. Agric. Forest.* 39, 515–530. doi: 10.3906/tar-1406-176

Jenkins, D., Dobert, R., Atanassova, A., and Pavely, C. (2021). Impacts of the regulatory environment for gene editing on delivering beneficial products. *In Vitro Cell. Dev. Biol. Plant* 57, 609–626. doi: 10.1007/s11627-021-10201-4

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), 1–4. doi: 10.1038/NPLANTS.2015.144

Jia, H., and Wang, N. (2020). Generation of homozygous canker-resistant citrus in the T0 generation using CRISPR-SpCas9p. *Plant Biotechnol. J.* 18 (10), 1990. doi: 10.1111/pbi.13375

Jia, H., Zhang, Y., Orbović, V., Xu, J., White, F. F., Jones, J. B., et al. (2017). Genome editing of the disease susceptibility gene Cs LOB 1 in citrus confers resistance to citrus canker. *Plant Biotechnol. J.* 15 (7), 817–823. doi: 10.1111/pbi.12677

Jiang, Y., Chai, Y. P., Lu, M. H., Han, X. L., Lin, Q., Zhang, Y., et al. (2020). Prime editing efficiently generates W542L and S621I double mutations in two ALS genes in maize. *Genome Biol.* 21, 257. doi: 10.1186/s13059-020-02170-5

Jones, R. A. C., and Barbetti, M. J. (2012). Influence of climate change on plant disease infections and epidemics caused by viruses and bacteria. *CAB Rev.: Perspect. Agric. Vet. Sci. Nutr. Nat. Resour.* 7, 1–31. doi: 10.1079/PAVSNNR20127022

Jones, J. D., and Dangl, J. L. (2006). The plant immune system. *Nature* 444 (7117), 323–329. doi: 10.1038/nature05286

Jones, J. D., Vance, R. E., and Dangl, J. L. (2016). Intracellular innate immune surveillance devices in plants and animals. *Science* 354 (6316), aaf6395. doi: 10.1126/science.aaf6395

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

Karavolias, N. G., Horner, W., Abugu, M. N., and Evanega, S. N. (2021). Application of gene editing for climate change in agriculture. *Front. Sustain Food Syst.* 5. doi: 10.3389/fsufs.2021.685801

Kavas, M., Mostafa, K., Secgin, Z., Yerlikaya, B. A., Yıldırım, K., and Gökdemir, G. (2023). Genome-wide analysis of duf221 domain-containing gene family in common bean and identification of its role on abiotic and phytohormone stress response. *Genet. Resour. Crop Evol.* 70 (1), 169–188. doi: 10.1007/s10722-022-01421-7

Kebert, M., Kostić, S., Zlatković, M., Stojnic, S., Čapelja, E., Zorić, M., et al. (2022). Ectomycorrhizal Fungi Modulate Biochemical Response against Powdery Mildew Disease in Quercus robur L. *Forests* 13 (9), 1491. doi: 10.3390/f13091491

Khan, Z., Khan, S. H., Ahmad, A., Aslam, S., Mubarik, M. S., and Khan, S. (2019). CRISPR/dCas9-mediated inhibition of replication of begomoviruses. *Int. J. Agric. Biol.* 21, 711–718. doi: 10.17957/IJAB/15.0948

Khattab, I. M., Fischer, J., Kaźmierczak, A., Thines, E., and Nick, P. (2023). Ferulic acid is a putative surrender signal to stimulate programmed cell death in grapevines after infection with Neofusicoccum parvum. *Plant Cell Environ.* 46 (1), 339–358. doi: 10.1111/pce.14468

Kieu, N. P., Lenman, M., Wang, E. S., Petersen, B. L., and Andreasson, E. (2021). Mutations introduced in susceptibility genes through CRISPR/Cas9 genome editing confer increased late blight resistance in potatoes. *Sci. Rep.* 11 (1), 1–12. doi: 10.1038/s41598-021-83972-w

Kim, Y. A., Moon, H., and Park, C. J. (2019). CRISPR/Cas9-targeted mutagenesis of Os8N3 in rice to confer resistance to *Xanthomonas oryzae* pv. oryzae. *Rice* 12 (1), 1–3. doi: 10.1186/s12284-019-0325-7

King, M., Altdorff, D., Li, P., Galagedara, L., Holden, J., and Unc, A. (2018). Northward shift of the agricultural climate zone under 21st-century global climate change. *Sci. Rep.* 8, 1–10. doi: 10.1038/s41598-018-26321-8

Kis, A., Hamar, É., Tholt, G., Bán, R., and Havelda, Z. (2019). Creating highly efficient resistance against wheat dwarf virus in barley by employing CRISPR/Cas9 system. *Plant Biotechnol. J.* 17, 1004–1006. doi: 10.1111/pbi.13077

Kocmánková, E., Trnka, M., Juroch, J., Dubrovský, M., Semerádová, D., Možný, M., et al. (2009). Impact of climate change on the occurrence and activity of harmful organisms. *Plant Prot. Sci.* 45, S48–S52. doi: 10.17221/2835-pps

Kohlen, W., Charnikhova, T., Lammers, M., Pollina, T., Tóth, P., Haider, I., et al. (2012). The tomato CAROTENOID CLEAVAGE DIOXYGENASE 8 (SlCCD8) regulates rhizosphere signaling, plant architecture and a_ects reproductive development through strigolactone biosynthesis. *New Phytol.* 196, 535–547. doi: 10.1111/j.1469-8137.2012.04265.x

Kuscu, C., Parlak, M., Tufan, T., Yang, J., Szlachta, K., Wei, X., et al. (2017). CRISPR-STOP: gene silencing through base-editing-induced nonsense mutations. *Nat. Methods* 14, 710–712. doi: 10.1038/nmeth.4327

Kyrou, K., Hammond, A. M., Galizi, R., Kranjc, N., Burt, A., Beaghton, A. K., et al. (2018). A CRISPR–Cas9 gene drive targeting doublesex causes complete population suppression in caged Anopheles Gambiae mosquitoes. *Nat. Biotechnol.* 36 (11), 1062–1066. doi: 10.1038/nbt.4245

Lakatos, L., Csorba, T., Pantaleo, V., Chapman, E. J., Carrington, J. C., Liu, Y. P., et al. (2006). Small RNA binding is a common strategy to suppress RNA silencing by several viral suppressors. *EMBO J.* 25, 2768–2780. doi: 10.1038/sj.emboj.7601164

Lanna-Filho, R. (2023). An overview of plant resistance to plant-pathogenic bacteria. *Trop. Plant Pathol.* 13, 1–7. doi: 10.1007/s40858-023-00560-1

Lapin, D., and Van den Ackerveken, G. (2013). Susceptibility to plant disease: More than a failure of host immunity. *Trends Plant Sci.* 18, 546–554. doi: 10.1016/j.tplants.2013.05.005

Larson, M. H., Gilbert, L. A., Wang, X., Lim, W. A., Weissman, J. S., and Qi, L. S. (2013). CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nat. Protoc.* 8, 2180–2196. doi: 10.1038/nprot.2013.132

Lema, M. (2021). Regulatory assessment of off-target changes and spurious DNA insertions in gene-edited organisms for agri-food use. *J. Regul. Sci.* 9, 1–15. doi: 10.21423/JRS-V09I1LEMA

Li, W., Deng, Y., Ning, Y., He, Z., and Wang, G. L. (2020). Exploiting broad-spectrum disease resistance in crops: from molecular dissection to breeding. *Annu. Rev. Plant Biol.* 71, 575–603. doi: 10.1146/annurev-arplant-010720-022215

Li, M. Y., Jiao, Y. T., Wang, Y. T., Zhang, N., Wang, B. B., Liu, R. Q., et al. (2020). CRISPR/Cas9-mediated VvPR4b editing decreases downy mildew resistance in grapevine (Vitis vinifera L.). *Hortic. Res.* 7, 149. doi: 10.1038/s41438-020-00371-4

Li, Y., Liang, J., Deng, B., Jiang, Y., Zhu, J., Chen, L., et al. (2023). Applications and prospects of CRISPR/Cas9-mediated base editing in plant breeding. *Curr. Issues Mol. Biol.* 45, 918–935. doi: 10.3390/cimb45020059

Li, Z., Zhang, D., Xiong, X., Yan, B., Xie, W., Sheen, J., et al. (2017). A potent Cas9-derived gene activator for plant and mammalian cells. *Nat. Plants* 3, 930. doi: 10.1038/s41477-017-0046-0

Liu, L., Sonbol, F. M., Huot, B., Gu, Y., Withers, J., Mwimba, M., et al. (2016). Salicylic acid receptors activate jasmonic acid signalling through a non-canonical pathway to promote effector-triggered immunity. *Nat. Commun.* 7, 1–10. doi: 10.1038/ncomms13099

Liu, H., Soyars, C. L., Li, J., Fei, Q., He, G., Peterson, B. A., et al. (2018). CRISPR/Cas9-mediated resistance to cauliflower mosaic virus. *Plant Direct.* 2, 1–9. doi: 10.1002/pld3.47

Lo Presti, L., Lanver, D., Schweizer, G., Tanaka, S., Liang, L., Tollot, M., et al. (2015). Fungal effectors and plant susceptibility. *Annu. Rev. Plant Biol.* 66, 513–545. doi: 10.1146/annurev-arplant-043014-114623

Lowder, L. G., Paul, J. W., and Qi, Y. (2017b). *Multiplexed transcriptional activation or repression in plants using CRISPR-dCas9-based systems Plant Gene Regulatory Networks* (New York, NY: Humana Press), 167–184.

Lowder, L. G., Zhou, J., Zhang, Y., Malzahn, A., Zhong, Z., Hsieh, T.-F., et al. (2017a). Robust transcriptional activation in plants using multiplexed CRISPR-Act2.0 and mTALE-Act systems. *Mol. Plant* 11, 245–256. doi: 10.1016/j.molp.2017.11.010

Luo, D., Huguet-Tapia, J. C., Raborn, R. T., White, F. F., Brendel, V. P., and Yang, B. (2021). The Xa7 resistance gene guards the rice susceptibility gene SWEET14 against exploitation by the bacterial blight pathogen. *Plant Commun.* 2, 100164. doi: 10.1016/j.xplc.2021.100164

Lykogianni, M., Bempelou, E., Karamaouna, F., and Aliferis, K. A. (2021). Do pesticides promote or hinder sustainability in agriculture? The challenge of sustainable use of pesticides in modern agriculture. *Sci. Total Environ.* 795, 148625. doi: 10.1016/j.scitotenv.2021.148625

Macovei, A., Sevilla, N. R., Cantos, C., Jonson, G. B., Slamet-Loedin, I., Čermák, T., et al. (2018). Novel alleles of rice eIF4G generated by CRISPR/Cas9-targeted mutagenesis confer resistance to Rice tungro spherical virus. *Plant Biotechnol. J.* 16 (11), 1918–1927. doi: 10.1111/pbi.12927

Mallory, A. C., Elmayan, T., and Vaucheret, H. (2008). MicroRNA maturation and action-the expanding roles of ARGONAUTEs. *Curr. Opin. Plant Biol.* 11, 560–566. doi: 10.1016/j.pbi.2008.06.008

Malnoy, M., Viola, R., Jung, M. H., Koo, O. J., Kim, S., Kim, J. S., et al. (2016). DNA-free genetically edited grapevine and apple protoplast using CRISPR/Cas9 ribonucleoproteins. *Front. Plant Sci.* 7, 1904. doi: 10.3389/fpls.2016.01904

McDonald, B. A. (2014). Using dynamic diversity to achieve durable disease resistance in agricultural ecosystems. *Trop. Plant Pathol.* 39, 191–196. doi: 10.1590/S1982-56762014000300001

Mehta, D., Stürchler, A., Anjanappa, R. B., Zaidi, S. S. E. A., Hirsch-Hoffmann, M., Gruissem, W., et al. (2019). Linking CRISPR-Cas9 interference in cassava to the evolution of editing-resistant geminiviruses. *Genome Biol.* 20, 1–10. doi: 10.1186/s13059-019-1678-3

Miladinović, D., Antunes, D., Yildirim, K., Bakhsh, A., Cvejić, S., Kondić-Špika, A., et al. (2021). Targeted plant improvement through genome editing: from laboratory to field. *Plant Cell Rep.* 40, 935–951. doi: 10.1007/s00299-020-02655-4

Moon, J. Y., and Park, J. M. (2016). Cross-talk in viral defense signaling in plants. *Front. Microbiol.* 7. doi: 10.3389/fmicb.2016.02068

Moradpour, M., and Abdulah, S. N. A. (2020). CRISPR/dCas9 platforms in plants: strategies and applications beyond genome editing. *Plant Biotechnol. J.* 18 (1), 32–44. doi: 10.1111/pbi.13232

Mukhtar, M. S. (2013). Engineering NLR immune receptors for broad-spectrum disease resistance. *Trends Plant Sci.* 18 (9), 469–472. doi: 10.1016/j.tplants.2013.08.005

Naegeli, H., Bresson, J. L., Dalmay, T., Dewhurst, I. C., Epstein, M. M., Firbank, L. G., et al. (2020). Applicability of the EFSA Opinion on site-directed nucleases type 3 for the safety assessment of plants developed using site-directed nucleases type 1 and 2 and oligonucleotide-directed mutagenesis. *EFSA J.* 18, 1–14. doi: 10.2903/j.efsa.2020.6299

Naidoo, S., Slippers, B., Plett, J. M., Coles, D., and Oates, C. N. (2019). The road to resistance in forest trees. *Front. Plant Sci.* 10, 273. doi: 10.3389/fpls.2019.00273

Nakahara, K. S., and Masuta, C. (2014). Interaction between viral RNA silencing suppressors and host factors in plant immunity. *Curr. Opin. Plant Biol.* 20, 88–95. doi: 10.1016/j.pbi.2014.05.004

Nekrasov, V., Wang, C., Win, J., Lanz, C., Weigel, D., and Kamoun, S. (2017). Rapid generation of a transgene-free powdery mildew resistant tomato by genome deletion. *Sci. Rep.* 7, 1–6. doi: 10.1038/s41598-017-00578-x

Niehl, A., and Heinlein, M. (2019). Perception of double-stranded RNA in plant antiviral immunity. *Mol. Plant Pathol.* 20, 1203–1210. doi: 10.1111/mpp.12798

Nnadi, N. E., and Carter, D. A. (2021). Climate change and the emergence of fungal pathogens. *PloS Pathog.* 17, 1–6. doi: 10.1371/journal.ppat.1009503

Okita, T. W., and Delseny, M. (2023). Genome editing in plants: new advances and applications in plant biology and agriculture. *Plant Sci.* 328, 111577. doi: 10.1016/j.plantsci.2022.111577

Oliva, R., Ji, C., Atienza-Grande, G., Huguet-Tapia, J. C., Perez-Quintero, A., Li, T., et al. (2019). Broad-spectrum resistance to bacterial blight in rice using genome editing. *Nat. Biotechnol.* 37 (11), 1344–1350. doi: 10.1038/s41587-019-0267-z

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

Paula de Toledo Thomazella, D., Brail, Q., Dahlbeck, D., and Staskawicz, B. (2016). CRISPR-Cas9 mediated mutagenesis of a DMR6 ortholog in tomato confers broad-spectrum disease resistance. *BioRxiv* 20, 064824. doi: 10.1101/064824

Peng, A., Chen, S., Lei, T., Xu, L., He, Y., Wu, L., et al. (2017). Engineering canker-resistant plants through CRISPR/Cas9-targeted editing of the susceptibility gene Cs LOB 1 promoter in citrus. *Plant Biotechnol. J.* 15 (12), 1509–1519. doi: 10.1111/pbi.12733

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, 578–589. doi: 10.1111/pbi.12284

Pieterse, C. M. J., Pierik, R., and van Wees, S. C. M. (2014). Different shades of JAZ during plant growth and defense. *New Phytol.* 204, 261–264. doi: 10.1111/nph.13029

Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P., et al. (2013). Repurposing CRISPR as an RNA-γuided platform for sequence-specific control of gene expression. *Cell* 152, 1173–1183. doi: 10.1016/j.cell.2013.02.022

Rato, C., Carvalho, M. F., Azevedo, C., and Oblessuc, P. R. (2021). Genome editing for resistance against plant pests and pathogens. *Transgenic Re.* 30, 427–459. doi: 10.1007/s11248-021-00262-x

Raupp, F. M., and Spring, O. (2013). New sesquiterpene lactones from sunflower root exudate as germination stimulants for Orobanche cumana. *J. Agri. Food Chem.* 61, 10481–10487. doi: 10.1021/jf402392e

Ren, B., Yan, F., Kuang, Y., Li, N., Zhang, D., Zhou, X., et al. (2018). Improved base editor for efficiently inducing genetic variations in rice with CRISPR/Cas9-guided hyperactive hAID mutant. *Mol. Plant* 11, 623–626. doi: 10.1016/j.molp.2018.01.005

Rostocks, N. (2021). Implications of the efsa scientific opinion on site directed nucleases 1 and 2 for risk assessment of genome-edited plants in the euEU. *Agronomy* 11, 1–12. doi: 10.3390/agronomy11030572

Roy, A., Zhai, Y., Ortiz, J., Neff, M., Mandal, B., Mukherjee, S. K., et al. (2019). Multiplexed editing of a begomovirus genome restricts escape mutant formation and disease development. *PloS One* 14 (10), e0223765. doi: 10.1371/journal.pone.0223765

Ruiz-Ferrer, V., and Voinnet, O. (2009). Roles of plant small RNAs in biotic stress responses. *Annu. Rev. Plant Biol.* 60, 485–510. doi: 10.1146/annurev.arplant.043008.092111

Saijo, Y., Loo, E. P., and Yasuda, S. (2018). Pattern recognition receptors and signaling in plant–microbe interactions. *Plant J.* 93 (4), 592–613. doi: 10.1111/tpj.13808

Santillán Martínez, M. I., Bracuto, V., Koseoglou, E., Appiano, M., Jacobsen, E., Visser, R. G., et al. (2020). CRISPR/Cas9-targeted mutagenesis of the tomato susceptibility gene PMR4 for resistance against powdery mildew. *BMC Plant Biol.* 20 (1), 284. doi: 10.1186/s12870-020-02497-y

Savary, S., Willocquet, L., Pethybridge, S. J., Esker, P., McRoberts, N., and Nelson, A. (2019). The global burden of pathogens and pests on major food crops. *Nat. Ecol. Evol.* 3 (3), 430–439. doi: 10.1038/s41559-018-0793-y

Seçgin, Z., Kavas, M., and Yildirim, K. (2021). Optimization of Agrobacterium-mediated transformation and regeneration for CRISPR/ Cas9 genome editing of commercial tomato cultivars. *Turkish J. Agric. Forest.* 45, 704–716. doi: 10.3906/tar-2009-49101

Secgin, Z., Uluisik, S., Yıldırım, K., Abdulla, M. F., Mostafa, K., and Kavas, M. (2022). Genome-wide identification of the aconitase gene family in tomato (Solanum lycopersicum) and CRISPR-based functional characterization of SlACO2 on male-sterility. *Int. J. Mol. Sci.* 23 (22), 13963. doi: 10.3390/ijms232213963

Shimatani, Z., Kashojiya, S., Takayama, M., Terada, R., Arazoe, T., Ishii, H., et al. (2017). Targeted base editing in rice and tomato using a CRISPR-Cas9 cytidine deaminase fusion. *Nat. Biotechnol.* 35, 441–443. doi: 10.1038/nbt.3833

Spring, O. (2021). Sesquiterpene lactones in sunflower oil. *LWT* 142), 111047. doi: 10.1016/j.lwt.2021.111047

Strange, R. N., and Scott, P. R. (2005). Plant disease: A threat to global food security. *Annu. Rev. Phytopathol.* 43, 83–116. doi: 10.1146/annurev.phyto.43.113004.133839

Strobbe, S., Wesana, J., van der Straeten, D., and De Steur, H. (2023). Public acceptance and stakeholderviews of gene edited foods: a global overview. *Trends Biotechnol.* 41, 6: 736–6: 740. doi: 10.1016/j.tibtech.2022.12.011

Sucu, S., Yağcı, A., and Yıldırım, K. (2018). Changes in morphological, physiological traits and enzyme activity of grafted and ungrafted grapevine rootstocks under drought stress. *Erwerbs-obstbau* 60 (2), 127–136. doi: 10.1007/s10341-017-0345-7

Tamura, Y., Osawa, T., Tabuchi, K., Yamasaki, K., Niiyama, T., Sudo, S., et al. (2022). Estimating plant–insect interactions under climate change with limited data. *Sci. Rep.* 12, 1–11. doi: 10.1038/s41598-022-14625-9

Tashkandi, M., Ali, Z., Aljedaani, F., Shami, A., and Mahfouz, M. M. (2018). Engineering resistance against Tomato yellow leaf curl virus via the CRISPR/Cas9 system in tomato. *Plant Signal Behav.* 13, 1–7. doi: 10.1080/15592324.2018.1525996

Thomma, B. P. H. J., Nürnberger, T., and Joosten, M. H. A. J. (2011). Of PAMPs and effectors: The blurred PTI-ETI dichotomy. *Plant Cell.* 23, 4–15. doi: 10.1105/tpc.110.082602

Tian, S., Jiang, L., Cui, X., Zhang, J., Guo, S., Li, M., et al. (2018). Engineering herbicide-resistant watermelon variety through CRISPR/Cas9-mediated base-editing. *Plant Cell Rep.* 37, 1353–1356. doi: 10.1007/s00299-018-2299-0

Tiwari, S. B., Belachew, A., Ma, S. F., Young, M., Ade, J., Shen, Y., et al. (2012). The EDLL motif: a potent plant transcriptional activation domain from AP2/ERF transcription factors. *Plant J.* 70, 855–865. doi: 10.1111/j.1365-313X.2012.04935.x

Tripathi, J. N., Ntui, V. O., Ron, M., Muiruri, S. K., Britt, A., and Tripathi, L. (2019). CRISPR/Cas9 editing of endogenous banana streak virus in the B genome of Musa spp. overcomes a major challenge in banana breeding. *Commun. Biol.* 2 (1), 46. doi: 10.1038/s42003-019-0288-7

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

Truniger, V., and Aranda, M. A. (2009). Recessive resistance to plant viruses. *Adv. Virus Res.* 75, 119–231. doi: 10.1016/S0065-3527(09)07504-6

Tuncel, A., and Qi, Y. (2022). CRISPR/Cas mediated genome editing in potato: Past achievements and future directions. *Plant Sci.* 325, 111474. doi: 10.1016/j.plantsci.2022.111474

Tyagi, S., Kumar, R., Kumar, V., Won, S. Y., and Shukla, P. (2020). Engineering disease resistant plants through CRISPR-Cas9 technology. *GM Crops Food.* 12, 125–144. doi: 10.1080/21645698.2020.1831729

Uehling, J., Deveau, A., and Paoletti, M. (2017). Do fungi have an innate immune response? An NLR-based comparison to plant and animal immune systems. *PloS Pathog.* 13, 1–8. doi: 10.1371/journal.ppat.1006578

Van Schie, C. C. N., and Takken, F. L. W. (2014). Susceptibility genes 101: How to be a good host. *Annu. Rev. Phytopathol.* 52, 551–581. doi: 10.1146/annurev-phyto-102313-045854

Vargas, W. A., Sanz Martín, J. M., Rech, G. E., Rivera, L. P., Benito, E. P., Díaz-Mínguez, J. M., et al. (2012). Plant defense mechanisms are activated during biotrophic and necrotrophic development of colletotricum graminicola in maize. *Plant Physiol.* 158, 1342–1358. doi: 10.1104/pp.111.190397

Vega-Muñoz, I., Duran-Flores, D., Fernández-Fernández, Á. D., Heyman, J., Ritter, A., and Stael, S. (2020). Breaking bad news: dynamic molecular mechanisms of wound response in plants. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.610445

Veillet, F., Perrot, L., Chauvin, L., Kermarrec, M. P., Guyon-Debast, A., Chauvin, J. E., et al. (2019). Transgene-free genome editing in tomato and potato plants using agrobacterium-mediated delivery of a CRISPR/Cas9 cytidine base editor. *Int. J. Mol. Sci.* 20, 402. doi: 10.3390/ijms20020402

Verlaan, M. G., Hutton, S. F., Ibrahem, R. M., Kormelink, R., Visser, R. G. F., Scott, J. W., et al. (2013). The tomato yellow leaf curl virus resistance genes Ty-1 and Ty-3 are allelic and code for DFDGD-class RNA-dependent RNA polymerases. *PloS Genet.* 9, e1003399. doi: 10.1371/journal.pgen.1003399

Vogel, J. T., Walter, M. H., Giavalisco, P., Lytovchenko, A., Kohlen, W., Charnikhova, T., et al. (2010). SlCCD7 controls strigolactone biosynthesis, shoot branching and mycorrhiza-induced apocarotenoid formation in tomato. *Plant J.* 61, 300–311. doi: 10.1111/j.1365-313X.2009.04056.x

Voigt, C. A. (2014). Callose-mediated resistance to pathogenic intruders in plant defense-related papillae. *Front. Plant Sci.* 5. doi: 10.3389/fpls.2014.00168

Vos, I. A., Moritz, L., Pieterse, C. M. J., and van Wees, S. C. M. (2015). Impact of hormonal crosstalk on plant resistance and fitness under multi-attacker conditions. *Front. Plant Sci.* 6. doi: 10.3389/fpls.2015.00639

Wakabayashi, T., Hamana, M., Mori, A., Akiyama, R., Ueno, K., Osakabe, K., et al. (2019). Direct conversion of carlactonoic acid to orobanchol by cytochrome P450 CYP722C in strigolactone biosynthesis. *Sci. Adv.* 5, eaax9067. doi: 10.1126/sciadv.aax9067

Wang, L., Chen, S., Peng, A., Xie, Z., He, Y., and Zou, X. (2019). CRISPR/Cas9-mediated editing of CsWRKY22 reduces susceptibility to Xanthomonas citri subsp. citri in Wanjincheng orange (Citrus sinensis (L.) Osbeck). *Plant Biotechnol. Rep.* 13, 501–510. doi: 10.1007/s11816-019-00556-x

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

Wang, L.-Y., Lin, S.-S., Hung, T.-H., Li, T.-K., Lin, N.-C., and Shen, T.-L. (2012). p126 protein can independently suppress local and systemic RNA silencing. *Mol. Plant-Microbe Interact.* 25, 648–657. doi: 10.1094/MPMI-06-11-0155

Wang, Y., Mostafa, S., Zeng, W., and Jin, B. (2021). Function and mechanism of Jasmonic acid in plant responses to abiotic and biotic stresses. *Int. J. Mol.* 22 (16), 8568. doi: 10.3390/ijms22168568

Wang, F., Wang, C., Liu, P., Lei, C., Hao, W., Gao, Y., et al. (2016). Enhanced rice blast resistance by CRISPR/ Cas9-Targeted mutagenesis of the ERF transcription factor gene OsERF922. *PloS One* 11, 1–18. doi: 10.1371/journal.pone.0154027

Wang, S., Zong, Y., Lin, Q., Zhang, H., Chai, Z., Zhang, D., et al. (2020). Precise, predictable multi-nucleotide deletions in rice and wheat using APOBEC-Cas9. *Nat. Biotechnol.* 38, 1460–1465. doi: 10.1038/s41587-020-0566-4

Wasternack, C., and Song, S. (2017). Jasmonates: Biosynthesis, metabolism, and signaling by proteins activating and repressing transcription. *J. Exp. Bot.* 68, 1303–1321. doi: 10.1093/jxb/erw443

Weßling, R., Epple, P., Altmann, S., He, Y., Yang, L., Henz, S. R., et al. (2014). Convergent targeting of a common host protein-network by pathogen effectors from three kingdoms of life. *Cell Host Microbe* 16 (3), 364–375. doi: 10.1016/j.chom.2014.08.004

Will, S., Vangheluwe, N., Krause, D., Fischer, A. R. H., Jorasch, P., Kohl, C. H., et al. (2023). Communicating about plant breeding and genome editing in plants: Assessment of European stakeholders, sources, channels and content. *Food Energy Security* 12, 415–421. doi: 10.1002/fes3.415

Win, J., Chaparro-Garcia, A., Belhaj, K., Saunders, D. G. O., Yoshida, K., Dong, S., et al. (2012). Effector biology of plant-associated organisms: Concepts and perspectives. *Cold Spring Harb. Symp. Quant. Biol.* 77, 235–247. doi: 10.1101/sqb.2012.77.015933

Wise, R. P., Moscou, M. J., Bogdanove, A. J., and Whitham, S. A. (2007). Transcript profiling in host-pathogen interactions. *Annu. Rev. Phytopathol.* 45, 329–369. doi: 10.1146/annurev.phyto.45.011107.143944

Xie, X., Yoneyama, K., and Yoneyama, K. (2010). The strigolactone story. *Annu. Rev. Phytopathol.* 48, 93–117. doi: 10.1146/annurev-phyto-073009-114453

Xu, Z., Xu, X., Gong, Q., Li, Z., Li, Y., Wang, S., et al. (2019). Engineering broad-spectrum bacterial blight resistance by simultaneously disrupting variable TALE-binding elements of multiple susceptibility genes in rice. *Mol. Plant* 12 (11), 1434–1446. doi: 10.1016/j.molp.2019.08.006

Yang, J., Duan, G., Li, C., Liu, L., Han, G., Zhang, Y., et al. (2019). The crosstalks between jasmonic acid and other plant hormone signaling highlight the involvement of jasmonic acid as a core component in plant response to biotic and abiotic stresses. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.01349

Yang, Y.-X., Jalal Ahammed, G., Wu, C., Fan, S., and Zhou, Y.-H. (2015). Crosstalk among jasmonate, salicylate and ethylene signaling pathways in plant disease and immune responses. *Curr. Protein Pept. Sci.* 16, 450–461. doi: 10.2174/1389203716666150330141638

Yildirim, K., Boylu, B., Atici, E., Kahraman, T., and S Akkaya, M. (2012). In Turkish wheat cultivars the resistance allele of LR34 is ineffective against leaf rust. *J. Plant Dis. Protect.* 119, 135–141. doi: 10.1007/BF03356432

Yin, K., Han, T., Xie, K., Zhao, J., Song, J., and Liu, Y. (2019). Engineer complete resistance to Cotton Leaf Curl Multan virus by the CRISPR/Cas9 system in *Nicotiana benthamiana*. *Phytopathol. Res.* 1, 1–9. doi: 10.1186/s42483-019-0017-7

Yin, K., and Qiu, J. L. (2019). Genome editing for plant disease resistance: applications and perspectives. *Philos. Trans. R. Soc. B.* 374 (1767), 20180322. doi: 10.1098/rstb.2018.0322

Yıldırım, K., Kavas, M., Kaya, R., Seçgin, 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

Yıldırım, K., Kavas, M., Küçük, İ. S., Seçgin, Z., and Saraç, Ç. 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 (7), 6515.

Yıldırım, K., and Kaya, Z. (2017). Gene regulation network behind drought escape, avoidance and tolerance strategies in black poplar (*Populus nigra* L.). *Plant Physiol. Biochem.* 115, 183–199. doi: 10.1016/j.plaphy.2017.03.020

Yu, Y., Gui, Y., Li, Z., Jiang, C., Guo, J., and Niu, D. (2022). Induced systemic resistance for improving plant immunity by beneficial microbes. *Plants* 11, 1–19. doi: 10.3390/plants11030386

Zafar, K., Khan, M. Z., Amin, I., Mukhtar, Z., Yasmin, S., Arif, M., et al. (2020). Precise CRISPR-Cas9 mediated genome editing in super basmati rice for resistance against bacterial blight by targeting the major susceptibility gene. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.00575

Zaidi, S. S. E. A., Mukhtar, M. S., and Mansoor, S. (2018). Genome editing: targeting susceptibility genes for plant disease resistance. *Trends Biotechnol.* 36, 898–906. doi: 10.1016/j.tibtech.2018.04.005

Zeng, X., Luo, Y., Vu, N. T. Q., Shen, S., Xia, K., and Zhang, M. (2020). CRISPR/Cas9-mediated mutation of OsSWEET14 in rice cv. Zhonghua11 confers resistance to *Xanthomonas oryzae* pv. oryzae without yield penalty. *BMC Plant Biol.* 20, 1–11. doi: 10.1186/s12870-020-02524-y

Zhan, X., Zhang, F., Zhong, Z., Chen, R., Wang, Y., Chang, L., et al. (2019). Generation of virus-resistant potato plants by RNA genome targeting. *Plant Biotechnol. J.* 17 (9), 1814–1822. doi: 10.1111/pbi.13102

Zhang, Y., Bai, Y., Wu, G., Zou, S., Chen, Y., Gao, C., et al. (2017). Simultaneous modification of three homoeologs of TaEDR1 by genome editing enhances powdery mildew resistance in wheat. *Plant J.* 91 (4), 714–724. doi: 10.1111/tpj.13599

Zhang, R., Chen, S., Meng, X., Chai, Z., Wang, D., Yuan, Y., et al. (2020). Generating broad-spectrum tolerance to ALS-inhibiting herbicides in rice by base editing. *Sci. China Life Sci* 64 (10), 1624–1633. doi: 10.1007/s11427-020-1800-5

Zhang, X., Dodds, P. N., and Bernoux, M. (2017). What do we know about NOD-like receptors in plant immunity? *Annu. Rev. Phytopathol.* 55, 205–229. doi: 10.1146/annurev-phyto-080516-035250

Zhang, Z., Ge, X., Luo, X., Wang, P., Fan, Q., Hu, G., et al. (2018). Simultaneous editing of two copies of Gh14-3-3d confers enhanced transgene-clean plant defense against Verticillium dahliae in allotetraploid upland cotton. *Front. Plant Sci.* 9, 842. doi: 10.3389/fpls.2018.00842

Zhang, T., Mudgett, M., Rambabu, R., Bradley Abramson, R., Xinhua, D., Todd, P. M., et al. (2022). Retraction Note: Selective inheritance of target genes from only one parent of sexually reproduced F1 progeny in Arabidopsis. *Nat. Commun.* 13, 3270. doi: 10.1038/s41467-022-31001-3

Zhang, Y., Iaffaldano, B., and Qi, Y. (2021). CRISPR ribonucleoprotein-mediated genetic engineering in plants. *Plant Commun.* 2 (2), 100168. doi: 10.1016/j.xplc.2021.100168

Zhang, Y., and Li, X. (2019). Salicylic acid: biosynthesis, perception, and contributions to plant immunity. *Curr. Opin. Plant Biol.* 50, 29–36. doi: 10.1016/j.pbi.2019.02.004

Zhang, M., Liu, Q., Yang, X., Xu, J., Liu, G., Yao, X., et al. (2020). CRISPR/Cas9-mediated mutagenesis of Clpsk1 in watermelon to confer resistance to *Fusarium oxysporum* f.sp. niveum. *Plant Cell Rep.* 39, 589–595. doi: 10.1007/s00299-020-02516-0

Zhang, T., Mudgett, M., Rambabu, R., Abramson, B., Dai, X., Michael, T. P., et al. (2021). Selective inheritance of target genes from only one parent of sexually reproduced F1 progeny in *Arabidopsis*. *Nat. Commun.* 12 (1), 3854. doi: 10.1038/s41467-021-24195-5

Zhang, T., Zhao, Y., Ye, J., Cao, X., Xu, C., Chen, B., et al. (2019). Establishing CRISPR/Cas13a immune system conferring RNA virus resistance in both dicot and monocot plants. *Plant Biotechnol. J.* 17, 1185–1187. doi: 10.1111/pbi.13095

Zhang, T., Zheng, Q., Yi, X., An, H., Zhao, Y., Ma, S., et al. (2018). Establishing RNA virus resistance in plants by harnessing CRISPR immune system. *Plant Biotechnol. J.* 16, 1415–1423. doi: 10.1111/pbi.12881

Zhu, S., Jeong, R. D., Lim, G. H., Yu, K., Wang, C., Chandra-Shekara, A. C., et al. (2013). Double-stranded RNA-binding protein 4 is required for resistance signaling against viral and bacterial pathogens. *Cell Rep.* 4, 1168–1184. doi: 10.1016/j.celrep.2013.08.018

Zong, Y., Wang, Y., Li, C., Zhang, R., Chen, K., Ran, Y., et al. (2017). Precise base editing in rice, wheat and maize with a Cas9-cytidine deaminase fusion. *Nat. Biotechnol.* 35, 483–440. doi: 10.1038/nbt.3811

Zou, X., Du, M., Liu, Y., Wu, L., Xu, L., Long, Q., et al. (2021). CsLOB1 regulates susceptibility to citrus canker through promoting cell proliferation in citrus. *Plant J.* 106 (4), 1039–1057. doi: 10.1111/tpj.15217

Zou, X., Long, J., Zhao, K., Peng, A., Chen, M., Long, Q., et al. (2019). Overexpressing GH3. 1 and GH3. 1L reduces susceptibility to Xanthomonas citri subsp. citri by repressing auxin signaling in citrus (*Citrus sinensis* Osbeck). *PloS One* 14 (12), e0220017. doi: 10.1371/journal.pone.0220017

frontiers | Frontiers in Plant Science

TYPE Review
PUBLISHED 29 November 2023
DOI 10.3389/fpls.2023.1268085

OPEN ACCESS

EDITED BY
Galal Bakr Anis,
Agricultural Research Center, Egypt

REVIEWED BY
Federica Zanetti,
Università di Bologna, Italy
Maliheh Esfahanian,
Carnegie Institution for Science,
United States

*CORRESPONDENCE
Yuhong Zhang
pzhangyh@nefu.edu.cn

RECEIVED 27 July 2023
ACCEPTED 15 November 2023
PUBLISHED 29 November 2023

CITATION
Ma J, Wang H and Zhang Y (2023) Research progress on the development of pennycress (*Thlaspi arvense* L.) as a new seed oil crop: a review.
Front. Plant Sci. 14:1268085.
doi: 10.3389/fpls.2023.1268085

Research progress on the development of pennycress (*Thlaspi arvense* L.) as a new seed oil crop: a review

Jianyu Ma[1,2], Haoyu Wang[1,2] and Yuhong Zhang[1,2]*

[1]Key Laboratory of Forest Plant Ecology, Ministry of Education, Northeast Forestry University, Harbin, China, [2]Heilongjiang Provincial Key Laboratory of Ecological Utilization of Forestry-Based Active Substances, Harbin, China

Compared with other crops, pennycress (*Thlaspi arvense* L.) is a niche emerging oil crop. In recent years, research on pennycress has been increasingly reflected in various directions. Pennycress belongs to the *Brassicaceae* family and was introduced from Eurasia to North America. It has been found worldwide as a cultivated plant and weed. In this paper, we review the advantages of pennycress as a supplementary model plant of *Arabidopsis thaliana*, oil and protein extraction technology, seed composition analysis based on metabolomics, germplasm resource development, growth, and ecological impact research, abiotic stress, fatty acid extraction optimization strategy, and other aspects of studies over recent years. The main research directions proposed for the future are as follows: (1) assemble the genome of pennycress to complete its entire genome data, (2) optimize the extraction process of pennycress as biodiesel, (3) analyze the molecular mechanism of the fatty acid synthesis pathway in pennycress, and (4) the functions of key genes corresponding to various adversity conditions of pennycress.

KEYWORDS

***Thlaspi arvense* L., supplementary model plant, gene, seed oil, fatty acid, germplasm development, abiotic stress**

1 Introduction

Energy crisis has become a global problem to be solved urgently. The promotion of industrial development and the exploitation of oil have led to the near depletion of oil resources. The ever-growing energy consumption and the associated environmental issues have urged for new sources of energy. With the development of science, technology, and industry, oil resources are increasingly limited. Finding alternative and environment-friendly oil resources to meet the needs of social development has become an urgent problem. Biodiesel is a kind of an environment-protective energy source. *Thlaspi arvense* L., as a new oil crop, has attracted much attention. *T. arvense* L., with alternative names of pennycress, field pennycress, stinkweed, bastard cress, field thlaspi, and ail sauvage, belongs to *Brassicaceae* and introduced to North America from Eurasia (Warwick et al., 2002). This cultivated plant species is also found as a

weed in other places of the world. Pennycress can be grown in different regions of the world to produce food, feed, and fuel (including renewable aviation fuel) and benefit the ecosystem by providing sustainable living coverage (Marks et al., 2021; Mousavi-Avval and Shah, 2021b; Keadle et al., 2023). An abundance of scientific evidence shows that mankind must achieve carbon neutrality by 2050 or even earlier to protect civilization and ecosystems from the devastating consequences of climate change. To achieve this, we must not only reduce and replace the use of fossil fuels, such as biofuels, but also need carbon capture and storage (CCS) to restore the carbon dioxide in the atmosphere to the current level.

Many genes replicated in *Arabidopsis* exist in the form of a single copy in mustard, which indicates that it will be easier to study their functions in mustard (McCormick, 2018). *T. arvense* is relatively easy to grow with a larger biomass than *Arabidopsis*, so it is more suitable for biochemistry development and other research (Figure 1). *T. arvense* can achieve 93.4% carbon conversion efficiency (CCE), which is much higher than other oilseeds (Sedbrook and Durrett, 2020). It has been estimated that once planted in half of the US Midwest Corn Belt, pennycress could fix 40 Mt of carbon and yield 9.8 billion liters of oil and 17 Mkg of seed meal each year (Fan et al., 2013).

At present, there is still a lack of in-depth research on pennycress. On the basis of existing research in recent years, this paper briefly reviewed the research on pennycress in the aspects of chromosome level detection, ecological development and distribution, abiotic stress, germplasm resource development, fatty acid composition, etc., to assist in the research on pennycress.

2 Pennycress is expected to become a supplementary model plant of *Arabidopsis*

Although *Arabidopsis thaliana* has been widely used as a model plant, it is not eligible as a model plant to study crop performance problems due to its short stature and lack of agronomic value. Over the past decade, numerous studies have demonstrated that pennycress can serve as an alternative model system analogous to *Arabidopsis* as it is desirable for both laboratory tests and large-scale experiments in the field. In recent years, a method similar to *Arabidopsis thaliana* for the genetic transformation of pennycress through *Agrobacterium*-mediated vacuum infiltration has been developed, which can produce 0.5% transformed seeds (McGinn et al., 2019). The *TaFAE1*-CRISPR-Cas9_Hyg vector was introduced into pennycress plants using the method described above to produce mutations in the *FATTY ACID ELONGATION1* (*FAE1*) gene, therefore producing an edible seed oil. After having been transformed with the gene that controls *Euonymus alatus* diacylglycerol acetyltransferase (*EaDAcT*) using the soybean glycinin promoter, the crop would be able to produce a novel drop-in fuel for diesel engines with low viscosity.

Numerous approaches are used to dissent the molecular networks of *T. arvense*. A research has sequenced, assembled, and annotated the transcriptome of pennycress (Dorn et al., 2013). Of these transcripts, 35% were most similar to an *A. thaliana* gene, and 74% were with top hits to the *Brassicaceae*. To generate a draft genome of *Thlaspi arvense* line MN106, a hybrid sequencing approach was used to generate 47 Gb of DNA sequencing reads from both the Illumina and PacBio platforms, which was annotated using the MAKER pipeline (Dorn et al., 2015). Compared with other *Brassicaceae* species, pennycress gene homologues revealed a high sequence global conservation that especially participated in glucosinolate biosynthesis, metabolism, and transport pathways. It was found that the peptides of *T. arvense* and five cruciferous plants including *Arabidopsis thaliana*, *Arabidopsis lyrata*, *Brassica rapa*, *Capsella rubella*, and *Eutrema salsugineum* were highly conserved through a combination comparative analysis, and *E. salsugineum* possessed the highest proportion of highly similar predicted peptides. EMS is used as a mutagen to induce *T. arvense* mutation, resulting in a pool of thousands of mutation genes, most of which are traceable in *Arabidopsis* (Chopra et al., 2018).

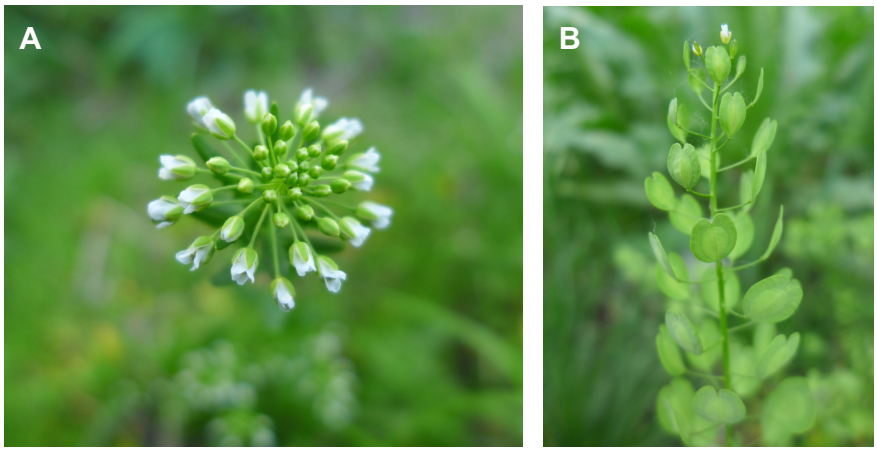


FIGURE 1
Pennycress growth morphology in the wild. **(A)** Raceme, petals white, spatulate, apex obtuse or emarginate. **(B)** Fruit obovate or suborbicular, base obtuse or rounded, apex deeply emarginate and apical notch.

The reduction of gene redundancy has promoted the identification of characters in *T. arvense* and the production of *T. arvense* mutants highly similar to *Arabidopsis* (Sedbrook et al., 2014).

A recent study suggested that the genome assembly of *T. arvense* var. MN106-Ref at the chromosome level with improved gene annotation can produce abundant variants, which represent both the genetic diversity in the collection and species population structure (Nunn et al., 2022). A sufficient variation in the transcripts was observed from two separate lines—MN108 and Spring32-10—during the analysis of transcriptome sequences. To avoid spurious associations and false positives, a unified mixed-model method is adopted, taking into account the population structure and kinship (Tandukar et al., 2022). It is the first report that specifically focuses on comprehending the genetic control of secondary domestication traits like seed size, oil content, and protein in pennycress populations in a multi-environment study. The results confirmed that genome-wide association study (GWAS) was an effective strategy to identify significant marker trait association, which was helpful for the breeding of pennycress.

3 Evaluation of the extraction technology of pennycress seed oil and protein

The pennycress seed contains 36% oil, which has erucic acid as its largest fatty acid (33%–38%), with linoleic and linolenic acids as the other two major components. (Moser et al., 2009b; Hojilla-Evangelista et al., 2013). It was found that pennycress oil could be converted into field pennycress oil methyl esters (FPME) (biodiesel) by a traditional alkali-catalyzed technique (Moser, 2012). In the assessment of the life cycle of pennycress jet fuel and diesel, a scenario analysis of the nitrogen fertilizer application amount, nitrogen content in crop residua, sources of H_2, and direct land use change (dLUC) has proved that the biofuels derived from *T. arvense* met the Renewable Fuel Standard (RFS2) and are promising to be widely used as advanced biofuels and biomass-based diesel (Fan et al., 2013; Moser et al., 2016; Mousavi-Avval and Shah, 2021a).

Compared with biodiesel extracted from commercial lipids, for instance, soybean, camelina, canola, sunflower, especially palm oils (Moser et al., 2016), the biodiesel produced has many splendid features, such as low cold filter plugging point (−17°C), cloud point temperature (−10°C), and pour point (−18°C) (Moser, 2012). Thereinto, the most important is the high cetane number and outstanding cold flow property (Moser et al., 2009b). Moser et al. (2009a) evaluated pennycress oil for the first time and evaluated that it conforms to the regulations of biodiesel according to the United States American Society for Testing and Materials (ASTM D6751). Nevertheless, the presence of longer-chain fatty acid methyl esters (FAME) has led to the kinematic viscosity of FPME being 5.24 mm^2/s at 40°C, which exceeds the 3.5–5.0 range specified in the European Committee for Standardization (CEN, EN 14214) (Moser, 2009; Moser et al., 2009a; Dunn, 2022). Therefore, for the purpose of better ameliorating the biodiesel characteristics of pennycress, it can be mixed with diesel oil with low kinematic viscosity to meet the EN 14214 (Moser et al., 2009a; Moser et al., 2016).

Isbell et al. (2015b) first concentrated erucic acid and methyl erucic acid by molecular distillation. The residence time of the molecular distillation unit is short under high temperatures, so it can distill macromolecular compound fatty acids (such as fatty acids or the corresponding methyl esters) without heat degradation like other distillation units, with rapid reaction and low cost (Isbell et al., 2015b).

The amino acid (AA) component of the protein in *T. arvense* is a typical plant protein, in which glycine, glutamic acid, and alanine are ubiquitous, with low-level essential amino acid content (Selling et al., 2013). The protein content produced by using 0.5 M sodium chloride at 5°C is the highest, reaching 83%, which is categorized to concentrate, while the extraction rate was only 25%. Pennycress crude protein extract had medium solubility (35%–45% soluble protein at pH 4), but it shows other excellent characteristics—foaming ability and foam stability (93%–97% residual foam) equalized with soybean protein and emulsification ability superior to lesquerella seed and the press cake (EAI > 150 m^2/g protein) (Hojilla-Evangelista et al., 2013; Hojilla-Evangelista et al., 2014). Thus, it is necessary to improve the extraction process to obtain a higher protein yield.

Two methods of protein extraction from pennycress (*T. arvense*) seed meal were evaluated, and the constituent, amino acid distribution, and chemical and physical properties (e.g., solvability, foamability, emulsifying, water holding capacity, and thermal coagulation) were compared with those of the synthesized protein (Hojilla-Evangelista et al., 2014; 2015). The results showed that the extraction method had an important effect on the purity and chemical and physical properties of the pressed protein products. Salt alkali precipitation (SE) was carried out through 0.1 mol/L NaCl at 50°C, while the traditional alkali dissolution and acid precipitation (AP) related to alkaline extraction (pH 10) were first followed by protein precipitation at pH 4. The crude protein extracted by SE and AP was at least 90% (db), which was the protein isolates (PI). A comparison of the results reveals that APPI had a lower protein yield (23%) but with a much higher purity (90% crude protein) than SE (45% yield and 67% crude protein) (Hojilla-Evangelista et al., 2014). Meanwhile, APPI showed higher foam capacity (120 ml), foam stability (96% foam volume retention), emulsification stability (24–35 min), and excellent heat resistance (solubility loss of 3% at pH 2 and pH 10). SEPI has better solubility (68%–91% at pH 2 and ≥7) and has a significantly competitive emulsifying activity than APPI (226–412 m^2/g protein). In short, SE and AP can extract protein isolates with outstanding characteristics from the seeds of *T. arvense*. Evangelista et al. (2012) determined that seed moisture content (MC) influences the pressing quality parameters of *T. arvense* seeds. Full-pressing and cooking have negative effects on the phosphatides and sulfur content in grains but have no effects on free fatty acid levels and oil color. The whole pennycress seed containing about 10% MC can be squeezed with minimum seed preparation conditions to obtain a press cake successfully, with an oil extraction rate of 10.7% (db) and 75.1% of the seed oil. Cooking and drying the seed MC between 3% and

4% offered a maximum oil recovery rate of 86.3% and 88.0%, respectively. The fast pyrolysis of defatted seed meal created stable high-carbon, low-oxygen (<30 wt%), high-energy liquid fuel intermediates that can manufacture jet fuels (Boateng et al., 2010; Evangelista et al., 2012). Transportation is one of the main cost reduction segments in the supply chain of producing sustainable aviation fuel (SAF) using pennycress (Trejo-Pech et al., 2019). A research has developed a GIS-based model that can provide suitable refinery site selection for SAF biorefineries based on pennycress (Mousavi-Avval et al., 2023). Although the final use of pennycress oil is mainly embodied in biodiesel, it is also planned to target the non-fuel industrial chemical marketplace and edible oil marketplace (Isbell et al., 2015a; Altendorf et al., 2019). Zhao et al. (2021) used the microwave-assisted biphasic extraction one-pot method for the first time to effectively extract oil and sinigrin from the seeds of pennycress.

4 Analysis of pennycress seed components based on metabolomics

The natural accumulation of erucic acid in *T. arvense* renders it as an outstanding oil crop and industrial crop. Metabolite fingerprints of the physiological activity of *T. arvense* embryos were analyzed by using non-targeted metabolomics and elective quantification of pivotal intracellular metabolites (more targeted) (Tsogtbaatar et al., 2015). A gas chromatography–mass spectrometry (GC–MS) analysis showed that there were three compound families of intracellular metabolites: organic acids (mainly malic acid and citric acid), amino acids, and sugar/sugar alcohols. Liquid chromatography tandem mass spectrometry (LC–MS/MS) was used to analyze the compounds obtained from boiling water extraction of different developmental stages of pennycress embryos. The results explored that glycolysis, oxidative pentose phosphate pathway, tricarboxylic acid cycle (TCA), and Calvin cycle were active in the growth process of pennycress.

According to the composition of liquid endosperm, Tsogtbaatar et al. (2020) established an *in vivo* culture system similar to the embryonic growth of apetala in plants. The biosynthetic efficiency of cultured pennycress embryos was measured as 93%, which was one of the highest compared with other oilseeds so far. The parallel labeling experiment with ^{13}C-labeled substrate in pennycress *T. arvense* revealed four reactions involved in fatty acid synthesis and extension. One is the oxidation reaction of the pentose phosphate pathway in the cytoplasm. The other is that isocitrate dehydrogenase (IDH), traditionally regarded as a catalytic and thermodynamic irreversible decarboxylation reaction, is reversible in pennycress embryos. Third, NADP-dependent malic enzyme (NADP-ME) generates pyruvate through the decarboxylation of malic acid transported in plastids. Fourth, Rubisco recycles carbon dioxide released by plastic pyruvate dehydrogenase (PHD) and malic enzyme, meanwhile providing a carbon skeleton for *de novo* petroleum synthesis. It should be noted that environmental impact should also be considered when evaluating breeding materials and predicting overall performance.

5 Development of germplasm resources of pennycress

Nowadays, the agricultural system in the Midwest of the USA is mainly a corn and soybean double-crop system (Johnson et al., 2017). The introduction of new species in this traditional agricultural system can bring higher economic and ecological benefits. During the fallow period, the potential production of *T. arvense* seeds is 1,120–2,240 kg · ha^{-1}, equivalent to 600–1,200 L · ha^{-1} of oil, while the potential yield of soybean and camellia seed oil is 450 and 420–640 L · ha^{-1}, respectively (Boateng et al., 2010; Phippen and Phippen, 2012). According to the 2-year (2009 to 2010) study by Phippen and Phippen (2012), the residue of pennycress in the field has no significant negative impact on the fatty acid content and other biomass of follow-up soybean crops. In terms of winter–spring production of fuel, *T. arvense* has considerable prospects. The experiments in the laboratory and the experimental plots proved that when 1.0 wt% of pennycress seed meal was integrated into the soil, it could effectively inhibit the germination and growth of weed seeds and would not replace the production of food crop and soybean (Isbell, 2009). Consequently, *T. arvense* can rotate with commercial crops without replacing them, and no extra land is required. Pennycress is known to exhibit both spring and winter types, sometimes within the same accession (Altendorf et al., 2019). Isbell et al. (2015a) reported that the seed germination rate of the first public nondormant pennycress winter line Katelyn (reg. no. GP-35, PI 673443) pennycress was 91% at immediate post-harvest, which was much higher than that of its original population Beecher (PI 672505) (only 7%). In spite of the fact that the germination rate of freshly harvested Katelyn (81%) seeds under light is much higher than that of the parent population (0%), it still does not meet the requirements of agricultural production (Isbell et al., 2017). Hence, an additional germplasm, Elizabeth (reg. no. GP-36, PI 677360), was developed (Isbell et al., 2017). According to the test, the germination rate of Elizabethan seeds in continuous darkness can reach 98% (Isbell et al., 2017). Elizabethan retains a winter type, the seeds need to be sowed after vernalization, and various mechanized methods are required for planting, such as unmanned aerial vehicles. Moreover, conventional herbicides can easily inhibit sprouted pennycress seeds (Isbell et al., 2015a). In another investigation, 41 winter type pennycress were collected from the United States Department of Agriculture (USDA) National Plant Germplasm System (NPGS) and wild. The yield rate of total oil, fatty acid distribution, and hundred seed weight were tested. It was found that although there were significant differences, there was no extreme variation, and mutation or additional germplasm resources were needed for domestication and improvement needed for breeding (Altendorf et al., 2019). A conserved CLAVTA3/ESR-related peptide family was identified in pennycress, including 27 gene members, most of which are involved in regulating biological processes such as root growth and shoot apical differentiation, regulating vascular bundle development, and affecting flowering and seed yield (Hagelthorn and Fletcher, 2023).

To sum up, finding novel traits and developing new breeding techniques for pennycress are necessary research directions in the future.

6 Research on growth and development and ecological impact

T. arvense belongs to the same family as *A. thaliana*, the *Brassicaceae*. Previous whole-genome sequencing (WGS) found that *T. arvense* was highly homologous with *A. thaliana*. Based on this, the following comparative genomics and other gene-level mining centers on whole-genome sequencing are widely used in the research on pennycress.

An interesting study is about the pennycress nectary, which is the first step to improving the nectar yield. Most *Brassicaceae* flowers have two pairs of non-equivalent nectaries; however, *T. arvense* flowers develop four equivalent nectaries. Between immature and mature nectaries of pennycress, over 3,000 genes were identified to be expressed differentially in gene ontology and metabolic pathway analyses (Thomas et al., 2017). The sugar yield of pennycress is much lower than camellia and canola, but it has a higher pollinator visit time (pvt) value because of its additional feed value to pollinating insects (Eberle et al., 2015). In addition, comparing the nectary transcriptome of pennycress and *Arabidopsis*, it is suggested that the mechanism of nectary maturation and nectary secretion of the two plants is highly conservative. Sucrose phosphate synthetase, which is necessary to produce nectar in *A. thaliana*, was not highly expressed in pennycress. Nevertheless, an *A. thaliana SUCROSE SYNTHASE1* (*SUS1*, Ta03482) homologous gene is profoundly expressed in mature pennycress nectaries. Therefore, Thomas et al. (2017) speculated that Ta03482 may replace the function of sucrose phosphate synthase in full-fledged pennycress nectaries.

Based on the results of comparative genomics, WGS, and co-segregation analysis, Dorn identified four natural alleles of FLC in pennycress that imply a spring annual growth habit (Dorn et al., 2018). The result suggested that mutation in FLC is the key factor leading to the loss of vernalization requirement in spring annual lines. Geng et al. (2021) *de novo* sequenced and assembled the genome of *T. arvense* from Kunming (southwest of China) at the chromosome level and detected genes related to DNA repair, ubiquitin system, and high-altitude adaptation. Notably, the *FLC* gene undergoes strong natural selection under a high altitude environment, and the functional loss of FLC protein caused by a single base mutation may cause premature flowering (Geng et al., 2021). The transcriptome analysis of *T. arvense* showed that the uplift of the Qinghai Tibet Plateau (QTP) had no significant impact on the population dynamics of *T. arvense*, which may be due to the rapid propagation of *T. arvense* seeds (An et al., 2015). However, natural selection in extreme habitats will prompt epigenetic differentiation among populations. *T. arvense* growing in extreme natural environmental conditions like low temperature, low oxygen, and high UV irradiation in the Tibetan Plateau of China has experienced strong natural selection; thus, the related genes have been highly differentiated (Guan et al., 2020). Hu et al. (2021) assembled a high-quality *T. arvense* genome and revealed the mechanism of plant response to extreme environments mediated by reverse transcriptional replication.

The RNA of 22 pennycress accessions derived from embryos at two developmental stages was analyzed using RNA-seq (García Navarrete et al., 2022). The data from this analysis support that the accession Ames 32872, originally from Armenia, is highly divergent from the other accessions, while the accessions originating from Canada and the United States cluster together.

The existence of highly similarity between the *Arabidopsis* and pennycress genomes has been determined by comparative genomics as well as the existence of a largely one-to-one correspondence between them. Using a shotgun mutagenesis approach, a *Ta-max3-like* dwarf mutant and *Ta-kcs5/cer60-like* wax mutant deficient in the long-chain fatty acid biosynthesis were also identified (Chopra et al., 2018). Methane has been recognized as the second greenhouse gas after carbon dioxide, which contributes to global warming (Keppler et al., 2006). In addition to anaerobic sources, plants can also emit CH_4 under aerobic conditions, which may harm the global methane budget. CH_4 emission mainly comes from plant vegetative parts and rarely from reproductive parts (Qaderi and Reid, 2014). In addition, stressed plants emit more methane than non-stressed plants, and the emission rate varies with plant species (Bruhn et al., 2012; Martel and Qaderi, 2019).

As an oilseed crop in the fallow period, pennycress appears in various crop rotation systems frequently, which will bring some economic benefits. A pennycress–soybean double-crop system can not only increase the total seed yield but also reduce the weed pressure in the cropping system (Johnson et al., 2015)—that is to say, pennycress stands established in autumn and will grow rapidly in spring to control early weed species (such as common lamb quarters and giant ragweed) and late weed species (such as tall water hemp) (Johnson et al., 2015). In the double-cropping system of *T. arvense* and camellia (*Camelina sativa* L. Crantz) in Germany, the flower pollination of *T. arvense* mainly depends on wind pollination, the contribution of self-pollination is only half of that of wind pollination, and the role of insect pollination is less (Groeneveld and Klein, 2014). Simultaneously, *T. arvense* provides abundant forage resources for different insect groups due to the fact that other plants rarely bloom in the same season.

To be honest, the growth of pennycress will also bring some other ecological changes. It is well known that there is a symbiotic relationship between *Thlaspi* and AMF fungi. Dean et al. (2017) characterized the fungi community, except AMF for pennycress, for the first time and found that the diversity and richness of the fungi community of soybean licking with pennycress as the covering crop were higher. In the corn–soybean cropping rotation system, pennycress is the substitute host of soybean cyst nematode (SCN; *Heterodera glycines*), whereas even at high incipient population density, SCN has no significant influence on the biomass of soybean and pennycress (Warwick et al., 2002; Hoerning et al., 2022).

7 Research on the abiotic stress of pennycress

Several studies have found that *T. arvense* responds to various abiotic stresses. Sharma et al. (2006) used microarray technology to discover that the gene response pattern of *T. arvense* under chilling stress remarkably resembles the model plant *Arabidopsis* and exhibits greater chilling tolerance compared to *A. thaliana* and *B. napus*. According to gene functional annotations and algorithms, 595 genes in *T. arvense* showed upregulation or downregulation under cold stress. The most striking observation is that the S-adenosylmethionine (AdoMet) gene is upregulated in the pathway controlling sulfate assimilation, which shows that increasing AdoMet may be an effective strategy for *T. arvense* to cope with cold stress. The CBF/DREB family is widely known to be involved in regulating cold stress-related transcription factors in various species. According to research, pennycress showed superior cold tolerance than other *Brassiceae* plants after 3 weeks of cold acclimation (Zhou et al., 2007). Moreover, the overexpression of the *TaCBF* gene identified in *T. arvense* enhanced the cold resistance of *Arabidopsis* (Zhou et al., 2007).

The influence of salinity (NaCl) stress on the epigenetic variation of DNA methylation in field *T. arvense* was investigated by using methylation-sensitive amplification polymorphism (MSAP) markers (Geng et al., 2020). Salt stress stimulation increased the apparent genetic diversity of the *T. arvense* population, and this stimulation could be partially transmitted to the offspring under a non-stress environment in the manner of stress memory. In addition, parallel changes in functional traits were observed, that is, the phenotypic variation of plants under salt stress was critically much higher than that of the control group.

In agreement with the findings of previous studies on *Arabidopsis*, it was observed upon extensive characterization of the *TRANSPARENT TESTA* 2 (*tt2*) mutant line that *tt2* can be susceptible to a number of abiotic stresses including osmotic stress, drought, and freezing (Chen et al., 2012; Ott et al., 2021).

The flow cytometric analyses of the leaves of two *Thlaspi* species suggested that *T. arvense* is a non-accumulator plant, while alpine pennycress (*Thlaspi caerulescens*) is a hyper-accumulator (Monteiro et al., 2010). No significant absorption or accumulation differences were detected in the determination of cadmium (Cd) and zinc (Zn) distribution in *T. arvense* leaves by laser ablation-inductively coupled plasma mass spectrometry (LA-ICP-MS) (Galiová et al., 2019). Since LA-ICP-MS is not sensitive to non-accumulator plants, it has been proved that pennycress is not super-accumulated to Cd and Zn. Further research found that Cd stress led to a decrease in water content, osmotic pressure, chlorophyll a and b content, and photosystem II efficiency in pennycress leaves and also led to nutrient imbalance in *T. arvense* roots (Monteiro and Soares, 2021).

Pennycress is sensitive to copper and is a suitable representative in understanding how common dicotyledons deal with middle-level metal pollution in urban ecosystems (Manceau et al., 2013). The analysis combined with synchronous X-ray fluorescence and absorption techniques showed that the concentration of copper in the roots of pennycress was 50–100 mg/kg, which exceeded the cell demand, accumulated in the cell wall of cortical and stellar root cells, and combined with nitrogen and oxygen donors in histidine residues (Manceau et al., 2013). Moreover, *T. arvense* is a super-rich selenium plant. Selenium polysaccharides with antioxidant and anti-tumor effects can be extracted from the leaves by efficient subcritical water extraction (SWE) (Xiang et al., 2022).

8 Extraction and content improvement strategy of pennycress seed oil and fatty acid

Pennycress seeds are composed of oil, protein, glucosinolates, tannins, fiber, and many other secondary metabolites, of which glucosinolates (about 100 μmol/g of seeds) were all in the form of sinigrin virtually (Warwick et al., 2002; Chopra et al., 2020a).

The high residue on the soil surface will also lead to the reduction of the establishment of grass–clover forests, which was observed in South Dakota in 2013 and 2014 (Eberle et al., 2015). Apart from genetic factors, the oleaginousness of field pennycress is also limited with environmental factors like soil temperatures and total rainfall (Cubins et al., 2019). Several research pointed that pennycress yield is closely related to soil temperature and humidity (Johnson et al., 2015; Zanetti et al., 2019). The oil accumulation of pennycress seeds is positively correlated with the vegetation period. Planting seedlings in the northern corn belt from late August to September will maximize the output and oil content (Dose et al., 2017). The higher accumulative rainfall, accumulative photohydrothermal period, and higher soil temperatures during sowing provide favorable environmental conditions for long-term growth at the cost of reducing the crude protein of the seeds so as to maximize the oil content (Dose et al., 2017). It was discovered that the seed thickness, 1,000-grain weight, oil content, peroxide value, acid value, and monounsaturated fatty acid content of *T. arvense* seeds were related to longitude and latitude, annual average rainfall, annual average temperature, and other geographical environments (Liu et al., 2022). Seed germination is also an important indicator for evaluating seed oil content. The total germination rate of seeds from different regions varies greatly, which has a significant impact on seed oil content and variety selection (Edo-Tena et al., 2020). In addition, Cubins et al. (2022) demonstrated that delaying the seed harvest time would cause only 26% of the loss, ensuring the maximum yield of pennycress seeds and the minimum loss of oil content at harvest time.

Same as the oil content, the erucic acid content was significantly different in different positions ($p < 0.05$), but it did not increase significantly through position interaction. Gesch et al. (2016) believed that there were variations in fatty acids in different locations, and they found that there were crucial differences in erucic acid in a single population in two regions. On the contrary, Claver et al. (2017) put forward the opposite view since there was no visible distinction in the oil characteristics of *T. arvense* seeds planted in the growth chamber and in field conditions.

As a newly developing oilseed crop, pennycress has irreplaceable advantages over other oilseed crops. Under the same

conditions, the growth cycle of pennycress is shorter than that of other oil crops, which is conducive to establishing a double-cropping system to replace winter fallow (Zanetti et al., 2019). However, differently from the conclusion of Dose et al. (2017); Zanetti et al. (2019) found that there was a strong association between the seed yield and the seed oil content of pennycress.

It is undeniable that the breeding of pennycress has just started and that it still has many characteristics that need to be improved. At present, *T. arvense* can be domesticated and engineered using a variety of molecular biological methods. The characteristics of weeds, such as seed dormancy and long survival time, are currently the target of domestication *via* CRISPR-Cas9 gene editing and mutation breeding techniques (Sedbrook et al., 2014). Screening and tendentious cultivation of these characters in pennycress will help to increase the net income and realize oilseed covering crops (Ott et al., 2019). The *TaAOP2* gene catalyzes the last step of the pennycress glucosinolate pathway. The mutation of this gene leads to the reduction of the metabolic flux of the pathway, but the specific mechanism is still unclear (Chopra et al., 2020b). The simultaneous mutation of both *TaROD1* (*REDUCED OLEATE DESATURATION1*) and *TaFAE1* (*FATTY ACID ELONGATION1*) genes can make the oleic acid content produced by pennycress much higher than that of the single mutants (Chopra et al., 2020a). The lipid profile of *T. arvense* with a double mutation of the *fad1* (*FATTY ACID DESATURASE2*) *rod1* gene is similar to those of canola seed TAGs, while the growth of plants with *fad2* single mutation and *fad2 fae1* double mutation was slow, and the overall height and seed yield decreased (Jarvis et al., 2021). More experiments are needed to explore the response of this genotype plant under various biological and abiotic stresses. The introduction of the acyltransferase (*LPAT* and *diacylglycerol acyltransferase*) gene and thioesterase (*FatB*) gene of its non-related species *Cuphea* into pennycress promoted the accumulation of a large amount of medium-chain fatty acids (MCFAs) in seeds without affecting the seed vitality (Esfahanian et al., 2021). According to the metabolomics and transcriptomics analysis of the 22 species of pennycress, improving the oil content of pennycress seeds can be achieved through four effective strategies: chlorophyll carbon allocation, gene modification related to lipid synthesis, enhancement of embryonic photosynthetic efficiency, and strict control of nitrogen utilization (Arias et al., 2023).

9 Conclusion

Pennycress is a promising new oilseed crop. There are many studies on the extraction technology of pennycress fuel. As an oil crop, it still has many characteristics that need further domestication and screening to increase its adaptability to the ecosystem. Recently, more scholars have contributed many excellent achievements to the research on pennycress species. Although several scholars have sequenced and assembled the genome of *T. arvense*, the function and various mechanisms of the *T. arvense* gene have not been fully explored. In particular, there is still a large gap in the research on pennycress in response to abiotic stress and the expression and regulation of related genes. Moreover, issues such as the standardized and unified planting standards for pennycress and the removal of sinigrin as proposed by Cubins et al. (2019) have not yet been resolved. There are many aspects to be solved in future research, such as the assembly and supplementation of genomes, combined with the analysis of the mechanism of regulating the fatty acid synthesis pathway in multi-omics, the cultivation of "zero erucic acid" pennycress seeds with healthy diet, or breeding the "high erucic acid" pennycress seeds for biofuels.

Author contributions

JM: Writing – original draft, Writing – review & editing. HW: Writing – review & editing. YZ: Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. The article supported by Heilongjiang Special Project Fund for the fourth Survey of Chinese Traditional Medicine Resources (NO: 2018HLJZYZYPC-16).

Conflict of interest

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

Publisher's note

References

Altendorf, K., Isbell, T., Wyse, D. L., and Anderson, J. A. (2019). Significant variation for seed oil content, fatty acid profile, and seed weight in natural populations of field pennycress (*Thlaspi arvense* L.). *Ind. Crops Products* 129, 261–268. doi: 10.1016/j.indcrop.2018.11.054

An, M., Zeng, L., Zhang, T., and Zhong, Y. (2015). Phylogeography of *Thlaspi arvense* (*Brassicaceae*) in China Inferred from chloroplast and nuclear DNA sequences and ecological niche modeling. *Int. J. Mol. Sci.* 16 (6), 13339–13355. doi: 10.3390/ijms160613339

Arias, C. L., García Navarrete, L. T., Mukundi, E., Swanson, T., Yang, F., Hernandez, J., et al. (2023). Metabolic and transcriptomic study of pennycress natural variation identifies targets for oil improvement. *Plant Biotechnol. J.* 21 (9), 1887–1903. doi: 10.1111/pbi.14101

Boateng, A. A., Mullen, C. A., and Goldberg, N. M. (2010). Producing stable pyrolysis liquids from the oil-Seed presscakes of mustard family plants: pennycress (*Thlaspi arvense* L.) and camelina (*Camelina sativa*). *Energy Fuels* 24 (12), 6624–6632. doi: 10.1021/ef101223a

Bruhn, D., Møller, I. M., Mikkelsen, T. N., and Ambus, P. (2012). Terrestrial plant methane production and emission. *Physiologia Plantarum* 144 (3), 201–209. doi: 10.1111/j.1399-3054.2011.01551.x

Chen, M., Wang, Z., Zhu, Y., Li, Z., Hussain, N., Xuan, L., et al. (2012). The effect of *TRANSPARENT TESTA2* on seed fatty acid biosynthesis and tolerance to environmental stresses during young seedling establishment in Arabidopsis. *Plant Physiol.* 160 (2), 1023–1036. doi: 10.1104/pp.112.202945

Chopra, R., Folstad, N., and Marks, M. D. (2020a). Combined genotype and fatty-acid analysis of single small field pennycress (*Thlaspi arvense*) seeds increases the throughput for functional genomics and mutant line selection. *Ind. Crops Products* 156, 112823. doi: 10.1016/j.indcrop.2020.112823

Chopra, R., Johnson, E. B., Daniels, E., McGinn, M., Dorn, K. M., Esfahanian, M., et al. (2018). Translational genomics using *Arabidopsis* as a model enables the characterization of pennycress genes through forward and reverse genetics. *Plant J.* 96 (6), 1093–1105. doi: 10.1111/tpj.14147

Chopra, R., Johnson, E. B., Emenecker, R., Cahoon, E. B., Lyons, J., Kliebenstein, D. J., et al. (2020b). Identification and stacking of crucial traits required for the domestication of pennycress. *Nat. Food* 1 (1), 84–91. doi: 10.1038/s43016-019-0007-z

Claver, A., Rey, R., López, M. V., Picorel, R., and Alfonso, M. (2017). Identification of target genes and processes involved in erucic acid accumulation during seed development in the biodiesel feedstock pennycress (*Thlaspi arvense* L.). *J. Plant Physiol.* 208, 7–16. doi: 10.1016/j.jplph.2016.10.011

Cubins, J. A., Wells, M. S., Frels, K., Ott, M. A., Forcella, F., Johnson, G. A., et al. (2019). Management of pennycress as a winter annual cash cover crop. A review. *Agron. Sustain. Dev.* 39 (5), 46. doi: 10.1007/s13593-019-0592-0

Cubins, J. A., Wells, S. S., Walia, M. K., Wyse, D. L., Becker, R., Forcella, F., et al. (2022). Harvest attributes and seed quality predict physiological maturity of pennycress. *Ind. Crops Products* 176, 114355. doi: 10.1016/j.indcrop.2021.114355

Dean, S. L., Billingsley Tobias, T., Phippen, W. B., Clayton, A. W., Gruver, J., and Porras-Alfaro, A. (2017). A study of *Glycine max* (soybean) fungal communities under different agricultural practices. *Plant Gene* 11, 8–16. doi: 10.1016/j.plgene.2016.11.003

Dorn, K. M., Fankhauser, J. D., Wyse, D. L., and Marks, M. D. (2013). *De novo* assembly of the pennycress (*Thlaspi arvense*) transcriptome provides tools for the development of a winter cover crop and biodiesel feedstock. *Plant J.* 75 (6), 1028–1038. doi: 10.1111/tpj.12267

Dorn, K. M., Fankhauser, J. D., Wyse, D. L., and Marks, M. D. (2015). A draft genome of field pennycress (*Thlaspi arvense*) provides tools for the domestication of a new winter biofuel crop. *DNA Res.* 22 (2), 121–131. doi: 10.1093/dnares/dsu045

Dorn, K. M., Johnson, E. B., Daniels, E. C., Wyse, D. L., and Marks, M. D. (2018). Spring flowering habit in field pennycress (*Thlaspi arvense*) has arisen multiple independent times. *Plant Direct* 2 (11), e00097. doi: 10.1002/pld3.97

Dose, H. L., Eberle, C. A., Forcella, F., and Gesch, R. W. (2017). Early planting dates maximize winter annual field pennycress (*Thlaspi arvense* L.) yield and oil content. *Ind. Crops Products* 97, 477–483. doi: 10.1016/j.indcrop.2016.12.039

Dunn, R. O. (2022). Fuel properties of low-erucic acid pennycress (LEAP) oil biodiesel. *Ind. Crops Products* 178, 114543. doi: 10.1016/j.indcrop.2022.114543

Eberle, C. A., Thom, M. D., Nemec, K. T., Forcella, F., Lundgren, J. G., Gesch, R. W., et al. (2015). Using pennycress, camelina, and canola cash cover crops to provision pollinators. *Ind. Crops Products* 75, 20–25. doi: 10.1016/j.indcrop.2015.06.026

Edo-Tena, E., Gesch, R. W., and Royo-Esnal, A. (2020). Germination patterns in seeds produced in apical and basal fruits of two *Thlaspi arvense* populations. *Agronomy* 10 (5), 756. doi: 10.3390/agronomy10050756

Esfahanian, M., Nazarenus, T. J., Freund, M. M., McIntosh, G., Phippen, W. B., Phippen, M. E., et al. (2021). Generating pennycress (*Thlaspi arvense*) seed triacylglycerols and acetyl-triacylglycerols containing medium-chain fatty acids. *Front. Energy Res.* 9. doi: 10.3389/fenrg.2021.620118

Evangelista, R. L., Isbell, T. A., and Cermak, S. C. (2012). Extraction of pennycress (*Thlaspi arvense* L.) seed oil by full pressing. *Ind. Crops Products* 37 (1), 76–81. doi: 10.1016/j.indcrop.2011.12.003

Fan, J., Shonnard, D. R., Kalnes, T. N., Johnsen, P. B., and Rao, S. (2013). A life cycle assessment of pennycress (*Thlaspi arvense* L.) -derived jet fuel and diesel. *Biomass Bioenergy* 55, 87–100. doi: 10.1016/j.biombioe.2012.12.040

Galiová, M. V., Száková, J., Prokeš, L., Čadková, Z., Coufalík, P., Kanický, V., et al. (2019). Variability of trace element distribution in Noccaea spp., *Arabidopsis* spp., and *Thlaspi arvense* leaves: the role of plant species and element accumulation ability. *Environ. Monit. Assess.* 191 (3), 181. doi: 10.1007/s10661-019-7331-5

García Navarrete, T., Arias, C., Mukundi, E., Alonso, A. P., and Grotewold, E. (2022). Natural variation and improved genome annotation of the emerging biofuel crop field pennycress (*Thlaspi arvense*). *G3 Genes|Genomes|Genetics* 12 (6), jkac084. doi: 10.1093/g3journal/jkac084

Geng, Y., Chang, N., Zhao, Y., Qin, X., Lu, S., Crabbe, M. J. C., et al. (2020). Increased epigenetic diversity and transient epigenetic memory in response to salinity stress in *Thlaspi arvense*. *Ecol. Evol.* 10 (20), 11622–11630. doi: 10.1002/ece3.6795

Geng, Y., Guan, Y., Qiong, L., Lu, S., An, M., Crabbe, M. J. C., et al. (2021). Genomic analysis of field pennycress (*Thlaspi arvense*) provides insights into mechanisms of adaptation to high elevation. *BMC Biol.* 19 (1), 143. doi: 10.1186/s12915-021-01079-0

Gesch, R. W., Royo-Esnal, A., Edo-Tena, E., Recasens, J., Isbell, T. A., and Forcella, F. (2016). Growth environment but not seed position on the parent plant affect seed germination of two *Thlaspi arvense* L. populations. *Ind. Crops Products* 84, 241–247. doi: 10.1016/j.indcrop.2016.02.006

Groeneveld, J. H., and Klein, A.-M. (2014). Pollination of two oil-producing plant species: Camelina (*Camelina sativa* L. Crantz) and pennycress (*Thlaspi arvense* L.) double-cropping in Germany. *GCB Bioenergy* 6 (3), 242–251. doi: 10.1111/gcbb.12122

Guan, Y., Qu, P., Lu, S., Crabbe, M. J. C., Zhang, T., and Geng, Y. (2020). Spatial genetic and epigenetic structure of *Thlaspi arvense* (field pennycress) in China. *Genes Genet. Syst.* 95 (5), 225–234. doi: 10.1266/ggs.20-00025

Hagelthorn, L., and Fletcher, J. C. (2023). The CLAVATA3/ESR-related peptide family in the biofuel crop pennycress. *Front. Plant Sci.* 14. doi: 10.3389/fpls.2023.1240342

Hoerning, C., Chen, S., Frels, K., Wyse, D., Wells, S., and Anderson, J. (2022). Soybean cyst nematode population development and its effect on pennycress in a greenhouse study. *J. Nematol.* 54 (1), 20220006. doi: 10.2478/jofnem-2022-0006

Hojilla-Evangelista, M. P., Evangelista, R. L., Isbell, T. A., and Selling, G. W. (2013). Effects of cold-pressing and seed cooking on functional properties of protein in pennycress (*Thlaspi arvense* L.) seed and press cakes. *Ind. Crops Products* 45, 223–229. doi: 10.1016/j.indcrop.2012.12.026

Hojilla-Evangelista, M. P., Selling, G. W., Berhow, M. A., and Evangelista, R. L. (2014). Preparation, composition and functional properties of pennycress (*Thlaspi arvense* L.) seed protein isolates. *Ind. Crops Products* 55, 173–179. doi: 10.1016/j.indcrop.2014.02.016

Hojilla-Evangelista, M. P., Selling, G. W., Berhow, M. A., and Evangelista, R. L. (2015). Extraction, composition and functional properties of pennycress (*Thlaspi arvense* L.) press cake protein. *J. Am. Oil Chemists' Soc.* 92 (6), 905–914. doi: 10.1007/s11746-015-2653-0

Hu, Y., Wu, X., Jin, G., Peng, J., Leng, R., Li, L., et al. (2021). Rapid genome evolution and adaptation of *Thlaspi arvense* mediated by recurrent RNA-Based and tandem gene duplications. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.772655

Isbell, T. A. (2009). US effort in the development of new crops (Lesquerella, Pennycress Coriander and Cuphea). *Oléagineux Corps gras Lipides* 16 (4-5-6), 205–210. doi: 10.1051/ocl.2009.0269

Isbell, T. A., Cermak, S. C., Dierig, D. A., Eller, F. J., and Marek, L. F. (2015a). Registration of Katelyn *Thlaspi arvense* L. (pennycress) with improved nondormant traits. *J. Plant Registrations* 9 (2), 212–215. doi: 10.3198/jpr2014.08.0053crg

Isbell, T. A., Cermak, S. C., and Marek, L. F. (2017). Registration of Elizabeth *Thlaspi arvense* L. (pennycress) with Improved nondormant traits. *J. Plant Registrations* 11 (3), 311–314. doi: 10.3198/jpr2016.12.0073crg

Isbell, T. A., Evangelista, R., Glenn, S. E., Devore, D. A., Moser, B. R., Cermak, S. C., et al. (2015b). Enrichment of erucic acid from pennycress (*Thlaspi arvense* L.) seed oil. *Ind. Crops Products* 66, 188–193. doi: 10.1016/j.indcrop.2014.12.050

Jarvis, B. A., Romsdahl, T. B., McGinn, M. G., Nazarenus, T. J., Cahoon, E. B., Chapman, K. D., et al. (2021). CRISPR/Cas9-Induced *fad2* and *rod1* mutations stacked with *fae1* confer high oleic acid seed oil in pennycress (*Thlaspi arvense* L.). *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.652319

Johnson, G. A., Kantar, M. B., Betts, K. J., and Wyse, D. L. (2015). Field pennycress production and weed control in a double crop system with soybean in Minnesota. *Agron. J.* 107 (2), 532–540. doi: 10.2134/agronj14.0292

Johnson, G. A., Wells, M. S., Anderson, K., Gesch, R. W., Forcella, F., and Wyse, D. L. (2017). Yield tradeoffs and nitrogen between pennycress, camelina, and soybean in relay- and double- crop systems. *Agron. J.* 109 (5), 2128–2135. doi: 10.2134/agronj2017.02.0065

Keadle, S. B., Sykes, V. R., Sams, C. E., Yin, X., Larson, J. A., and Grant, J. F. (2023). National winter oilseeds review for potential in the US Mid-South: Pennycress, Canola, and Camelina. *Agron. J.* 115 (3), 1415–1430. doi: 10.1002/agj2.21317

Keppler, F., Hamilton, J. T., Brass, M., and Röckmann, T. (2006). Methane emissions from terrestrial plants under aerobic conditions. *Nature* 439 (7073), 187–191. doi: 10.1038/nature04420

Liu, J., Chen, M., Zhang, Y., and Zheng, B. (2022). Analyses of the oil content, fatty acid composition, and antioxidant activity in seeds of *Thlaspi arvense* L. from different provenances and correlations with environmental factors. *Chem. Biol. Technol. Agric.* 9 (1), 11. doi: 10.1186/s40538-021-00276-x

Manceau, A., Simionovici, A., Lanson, M., Perrin, J., Tucoulou, R., Bohic, S., et al. (2013). *Thlaspi arvense* binds Cu(II) as a bis-(L-histidinato) complex on root cell walls in an urban ecosystem. *Metallomics: Integrated Biometal Sci.* 5 (12), 1674–1684. doi: 10.1039/c3mt00215b

Marks, M. D., Chopra, R., and Sedbrook, J. C. (2021). Technologies enabling rapid crop improvements for sustainable agriculture: example pennycress (*Thlaspi arvense* L.). *Emerg. Topics Life Sci.* 5 (2), 325–335. doi: 10.1042/ETLS20200330

Martel, A. B., and Qaderi, M. M. (2019). Unravelling the effects of blue light on aerobic methane emissions from canola. *J. Plant Physiol.* 233, 12–19. doi: 10.1016/j.jplph.2018.12.006

McCormick, S. (2018). Ta Ta for now: *Thlapsi arvense* (pennycress), an emerging model for genetic analyses. *Plant J.* 96 (6), 1091–1092. doi: 10.1111/tpj.14172

McGinn, M., Phippen, W. B., Chopra, R., Bansal, S., Jarvis, B. A., Phippen, M. E., et al. (2019). Molecular tools enabling pennycress (*Thlaspi arvense*) as a model plant and oilseed cash cover crop. *Plant Biotechnol. J.* 17 (4), 776–788. doi: 10.1111/pbi.13014

Monteiro, M. S., Rodriguez, E., Loureiro, J., Mann, R. M., Soares, A. M. V. M., and Santos, C. (2010). Flow cytometric assessment of Cd genotoxicity in three plants with different metal accumulation and detoxification capacities. *Ecotoxicol. Environ. Saf.* 73 (6), 1231–1237. doi: 10.1016/j.ecoenv.2010.06.020

Monteiro, M. S., and Soares, A. M. V. M. (2021). Physiological and biochemical effects of Cd stress in *Thlaspi arvense* L—A non-accumulator of metals. *Arch. Environ. Contamination Toxicol.* 81 (2), 285–292. doi: 10.1007/s00244-021-00873-9

Moser, B. R. (2009). Biodiesel production, properties, and feedstocks. *In Vitro Cell. Dev. Biol. — Plant* 45, 229–266. doi: 10.1039/c3mt00215b

Moser, B. R. (2012). Biodiesel from alternative oilseed feedstocks: camelina and field pennycress. *Biofuels* 3 (2), 193–209. doi: 10.4155/bfs.12.6

Moser, B. R., Evangelista, R. L., and Isbell, T. A. (2016). Preparation and fuel properties of field pennycress (*Thlaspi arvense*) seed oil ethyl esters and blends with ultralow-sulfur diesel fuel. *Energy Fuels* 30 (1), 473–479. doi: 10.1021/acs.energyfuels.5b02591

Moser, B. R., Knothe, G., Vaughn, S. F., and Isbell, T. A. (2009a). Production and evaluation of biodiesel from field pennycress (*Thlaspi arvense* L.) oil. *Energy Fuels* 23 (8), 4149–4155. doi: 10.1021/ef900337g

Moser, B. R., Shah, S. N., Winkler-Moser, J. K., Vaughn, S. F., and Evangelista, R. L. (2009b). Composition and physical properties of cress (*Lepidium sativum* L.) and field pennycress (*Thlaspi arvense* L.) oils. *Ind. Crops Products* 30 (2), 199–205. doi: 10.1016/j.indcrop.2009.03.007

Mousavi-Avval, S. H., Khanal, S., and Shah, A. (2023). Assessment of potential pennycress availability and suitable sites for sustainable aviation fuel refineries in Ohio. *Sustainability* 15 (13), 10589. doi: 10.3390/su151310589

Mousavi-Avval, S. H., and Shah, A. (2021a). Life cycle energy and environmental impacts of hydroprocessed renewable jet fuel production from pennycress. *Appl. Energy* 297, 117098. doi: 10.1016/j.apenergy.2021.117098

Mousavi-Avval, S. H., and Shah, A. (2021b). Techno-economic analysis of hydroprocessed renewable jet fuel production from pennycress oilseed. *Renewable Sustain. Energy Rev.* 149, 111340. doi: 10.1016/j.rser.2021.111340

Nunn, A., Rodríguez-Arévalo, I., Tandukar, Z., Frels, K., Contreras-Garrido, A., Carbonell-Bejerano, P., et al. (2022). Chromosome-level *Thlaspi arvense* genome provides new tools for translational research and for a newly domesticated cash cover crop of the cooler climates. *Plant Biotechnol. J.* 20 (5), 944–963. doi: 10.1111/pbi.13775

Ott, M. A., Eberle, C. A., Thom, M. D., Archer, D. W., Forcella, F., Gesch, R. W., et al. (2019). Economics and agronomics of relay-cropping pennycress and camelina with soybean in Minnesota. *Agron. J.* 111 (3), 1281–1292. doi: 10.2134/AGRONJ2018.04.0277

Ott, M. A., Gardner, G., Rai, K. M., Wyse, D. L., Marks, M. D., and Chopra, R. (2021). TRANSPARENT TESTA 2 allele confers major reduction in pennycress (*Thlaspi arvense* L.) seed dormancy. *Ind. Crops Products* 174, 114216. doi: 10.1016/j.indcrop.2021.114216

Phippen, W. B., and Phippen, M. E. (2012). Soybean seed yield and quality as a response to field pennycress residue. *Crop Sci.* 52 (6), 2767–2773. doi: 10.2135/cropsci2012.03.0192

Qaderi, M. M., and Reid, D. M. (2014). Aerobic methane emissions from stinkweed (*Thlaspi arvense*) capsules. *Plant Signaling Behav.* 9 (10), e970095. doi: 10.4161/15592316.2014.970095

Sedbrook, J. C., and Durrett, T. P. (2020). Pennycress, carbon wise: labeling experiments reveal how pennycress seeds efficiently incorporate carbon into biomass. *J. Exp. Bot.* 71 (10), 2842–2846. doi: 10.1093/jxb/eraa136

Sedbrook, J. C., Phippen, W. B., and Marks, M. D. (2014). New approaches to facilitate rapid domestication of a wild plant to an oilseed crop: Example pennycress (*Thlaspi arvense* L.). *Plant Sci.* 227, 122–132. doi: 10.1016/j.plantsci.2014.07.008

Selling, G. W., Hojilla-Evangelista, M. P., Evangelista, R. L., Isbell, T., Price, N., and Doll, K. M. (2013). Extraction of proteins from pennycress seeds and press cake. *Ind. Crops Products* 41, 113–119. doi: 10.1016/j.indcrop.2012.04.009

Sharma, N., Cram, D., Huebert, T., Zhou, N., and Parkin, I. A. P. (2006). Exploiting the wild crucifer *Thlaspi arvense* to identify conserved and novel genes expressed during a plant's response to cold stress. *Plant Mol. Biol.* 63 (2), 171–184. doi: 10.1007/s11103-006-9080-4

Tandukar, Z., Chopra, R., Frels, K., Heim, B., Marks, M. D., and Anderson, J. A. (2022). Genetic dissection of seed characteristics in field pennycress *via* genome-wide association mapping studies. *Plant Genome* 15 (2), e20211. doi: 10.1002/tpg2.20211

Thomas, J. B., Hampton, M. E., Dorn, K. M., David Marks, M., and Carter, C. J. (2017). The pennycress (*Thlaspi arvense* L.) nectary: structural and transcriptomic characterization. *BMC Plant Biol.* 17 (1), 201. doi: 10.1186/s12870-017-1146-8

Trejo-Pech, C. O., Larson, J. A., English, B. C., and Yu, T. E. (2019). Cost and profitability analysis of a prospective pennycress to sustainable aviation fuel supply chain in southern USA. *Energies* 12 (16), 3055. doi: 10.3390/en12163055

Tsogtbaatar, E., Cocuron, J.-C., and Alonso, A. P. (2020). Non-conventional pathways enable pennycress (*Thlaspi arvense* L.) embryos to achieve high efficiency of oil biosynthesis. *J. Exp. Bot.* 71 (10), 3037–3051. doi: 10.1093/jxb/eraa060

Tsogtbaatar, E., Cocuron, J.-C., Sonera, M. C., and Alonso, A. P. (2015). Metabolite fingerprinting of pennycress (*Thlaspi arvense* L.) embryos to assess active pathways during oil synthesis. *J. Exp. Bot.* 66 (14), 4267–4277. doi: 10.1093/jxb/erv020

Warwick, S. I., Francis, A., and Susko, D. J. (2002). The biology of Canadian weeds. 9. *Thlaspi arvense* L. (updated). *Can. J. Plant Sci.* 82 (4), 803–823. doi: 10.4141/P01-159

Xiang, A., Li, W., Zhao, Y., Ju, H., Xu, S., Zhao, S., et al. (2022). Purification, characterization and antioxidant activity of selenium-containing polysaccharides from pennycress (*Thlaspi arvense* L.). *Carbohydr. Res.* 512, 108498. doi: 10.1016/j.carres.2021.108498

Zanetti, F., Isbell, T. A., Gesch, R. W., Evangelista, R. L., Alexopoulou, E., Moser, B., et al. (2019). Turning a burden into an opportunity: Pennycress (*Thlaspi arvense* L.) a new oilseed crop for biofuel production. *Biomass Bioenergy* 130, 105354. doi: 10.1016/j.biombioe.2019.105354

Zhao, R., Wei, M., Shi, G., Wang, X., Gao, H., Zhang, L., et al. (2021). One-pot process for simultaneously obtaining oil and sinigrin from field pennycress (*Thlaspi arvense*) seeds using microwave-assisted biphasic extraction. *Ind. Crops Products* 166, 113483. doi: 10.1016/j.indcrop.2021.113483

Zhou, N., Robinson, S. J., Huebert, T., Bate, N. J., and Parkin, I. A. P. (2007). Comparative genome organization reveals a single copy of *CBF* in the freezing tolerant crucifer *Thlaspi arvense*. *Plant Mol. Biol.* 65 (5), 693–705. doi: 10.1007/s11103-007-9235-y

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 06 December 2023
DOI 10.3389/fpls.2023.1288997

OPEN ACCESS

EDITED BY
Galal Bakr Anis,
Agricultural Research Center, Egypt

REVIEWED BY
Ahmed Ramadn,
King Abdulaziz University, Saudi Arabia
Muhammad Zeeshan Mansha,
University of Layyah, Pakistan
Umair Ashraf,
University of Education Lahore, Pakistan

*CORRESPONDENCE
Ning Lv
lvn@gsau.edu.cn

RECEIVED 05 September 2023
ACCEPTED 20 November 2023
PUBLISHED 06 December 2023

CITATION
Liu H-p, Yang Q-y, Liu J-x, Haq IU, Li Y, Zhang Q-y, Attia KA, Abushady AM, Liu C-z and Lv N (2023) Host plant-mediated effects on *Buchnera* symbiont: implications for biological characteristics and nutritional metabolism of pea aphids (*Acyrthosiphon pisum*). *Front. Plant Sci.* 14:1288997. doi: 10.3389/fpls.2023.1288997

Host plant-mediated effects on *Buchnera* symbiont: implications for biological characteristics and nutritional metabolism of pea aphids (*Acyrthosiphon pisum*)

Hui-ping Liu[1], Qiao-yan Yang[1], Jing-xing Liu[1], Inzamam Ul Haq[1], Yan Li[1], Qiang-yan Zhang[1], Kotb A. Attia[2], Asmaa M. Abushady[3,4], Chang-zhong Liu[1] and Ning Lv[1]*

[1]Biocontrol Engineering Laboratory of Crop Diseases and Pests of Gansu Province, College of Plant Protection, Gansu Agricultural University, Lanzhou, China, [2]Department of Biochemistry, College of Science, King Saud University, Riyadh, Saudi Arabia, [3]Biotechnology School, Nile University, 26th of July Corridor, Sheikh Zayed City, Giza, Egypt, [4]Department of Genetics, Agriculture College, Ain Shams University, Cairo, Egypt

Introduction: The pea aphid, *Acyrthosiphon pisum*, is a typical sap-feeding insect and an important worldwide pest. There is a primary symbiont-*Buchnera aphidicola*, which can synthesize and provide some essential nutrients for its host. At the same time, the hosts also can actively adjust the density of bacterial symbiosis to cope with the changes in environmental and physiological factors. However, it is still unclear how symbionts mediate the interaction between herbivorous insects' nutrient metabolism and host plants.

Methods: The current study has studied the effects of different host plants on the biological characteristics, *Buchnera* titer, and nutritional metabolism of pea aphids. This study investigated the influence of different host plants on biological characteristics, *Buchnera* titer, and nutritional metabolism of pea aphids.

Results and discussion: The titer of *Buchnera* was significantly higher on *T. Pretense* and *M. officinalis*, and the relative expression levels were 1.966±0.104 and 1.621±0.167, respectively. The content of soluble sugar (53.46±1.97μg/mg), glycogen (1.12±0.07μg/mg) and total energy (1341.51±39.37μg/mg) of the pea aphid on *V. faba* were significantly higher and showed high fecundity (143.86 ±11.31) and weight (10.46±0.77μg/mg). The content of total lipids was higher on *P. sativum* and *T. pretense*, which were 2.82±0.03μg/mg and 2.92±0.07μg/mg, respectively. Correlation analysis found that the difference in *Buchnera* titer was positively correlated with the protein content in *M. officinalis* and the content of total energy in *T. pratense* ($P < 0.05$). This study confirmed that host plants not only affected the biological characteristics and nutritional metabolism of pea aphids but also regulated the symbiotic density, thus interfering with the nutritional function of *Buchnera*. The results can provide a theoretical basis for further studies on the influence of different host plants on the development of pea aphids and other insects.

KEYWORDS

host plant, *Acyrthosiphon pisum*, primary symbiont, development and growth, nutrients

1 Introduction

Symbiosis, a complex and mutually advantageous relationship between species, has consistently been a fundamental aspect of biological accomplishments (Vostinar et al., 2021). Within the expansive and heterogeneous realm of arthropods, a notable illustration of such cooperative associations can be observed in their interactions with microbes, specifically endosymbiotic bacteria (Perlmutter and Bordenstein, 2020). A remarkable finding that arises from current research is the prevalence and importance of symbiotic relationships in arthropods, with over 50% of arthropod species engaging in some partnership with symbionts (Weinert et al., 2015; Harumoto et al., 2016). One can witness the epitome of mutualistic relationships by delving further into the domain of Hemiptera, particularly the sub-order Sternorrhyncha, which encompasses psyllids, aphids, and mealybugs. These hemipteran insects engage in obligatory and facultative associations with various prokaryotic endosymbionts, rendering them a valuable resource for researchers seeking to comprehend the complexities of symbiotic partnerships (Douglas, 1998; Baumann, 2005).

The obligate symbionts, also known as primary symbionts, are notable for their essential function in the existence of their host insects. The role of these organisms extends beyond mere cohabitation, as they provide their hosts with vital nutrients, including amino acids and vitamins, which are typically lacking in their primary diet consisting of plant sap. The mutualistic association between insects and their endosymbiotic bacteria, such as the pea aphid-*Buchnera* and whitefly-*Portiera* models, has been extensively studied and well-documented in scientific literature. These studies have provided comprehensive insights into the essential functions performed by these bacteria in ensuring the survival of their host organisms (Sloan et al., 2014; Douglas, 2015; Morrow et al., 2017; Bao et al., 2021). At a more intricate molecular level, these obligate symbionts effectively enhance metabolic pathways within their hosts, thereby compensating for any nutritional deficiencies. The examination of this specific symbiotic alliance is of notable significance, as it provides insight into the mechanisms by which nature ensures survival in the face of dietary constraints (Wilson and Duncan, 2015; Bao et al., 2021; Zhu et al., 2022). One illustrative instance is the symbiotic relationship between *Francisella*, a microorganism, and the African soft tick, *Ornithodoros moubata*. This symbiosis is characterized by the production of vital molecules such as biotin and riboflavin by *Francisella*, which enhances the nutritional value of the tick's diet. This example highlights the interdependence between the host organism and its symbiotic partner (Duron et al., 2018).

In contrast, secondary endosymbionts, while not indispensable for the viability of their hosts, play significant roles in influencing their ecological dynamics and behavioral patterns. The studies mentioned above by Lv et al. (2021) and Deehan et al. (2021) have demonstrated that microorganisms can exert an impact on dietary preferences, enhance defense mechanisms against predators and pathogens, and potentially induce modifications in reproductive strategies. Himler et al. (2011) reported an intriguing finding regarding Rickettsia-infected whiteflies, wherein the presence of the symbiont resulted in a range of advantages, including enhanced progeny production and a biased sex ratio favoring females. Moreover, the exceptional capabilities of secondary endosymbionts are demonstrated by their capacity to induce distinctive reproductive consequences, such as parthenogenesis and male killing. These intriguing phenomena have captured the interest of evolutionary biologists (Stouthamer et al., 1999; Werren et al., 2008).

The intricate interplay between hosts and their symbionts, characterized by genomic and metabolic integration, is a crucial aspect that should not be disregarded, as it has undergone significant refinement throughout evolutionary timescales (Moran et al., 2008; Zhao et al., 2020). The co-evolutionary trajectories and genome compatibility of the organisms under study are clearly indicative of a profound level of integration, as demonstrated by Wilson and Duncan (2015) and Moran and Bennett (2014). It is worth noting that the titers of endosymbionts within their respective hosts are not fixed entities. They exhibit dynamic adjustments in response to a wide range of external and internal factors. Various factors can influence the densities of endosymbionts, including the genetic composition of the host, environmental stressors, alterations in diet, and exposure to antibiotics (Cassone et al., 2015; Zhang et al., 2016; Nguyen et al., 2017; Zhang et al., 2019; Bulman et al., 2021). As an illustration, the investigation conducted by Patton et al. (2021) revealed that specific viral infections in aphids resulted in a significant decrease in the densities of their *Buchnera* symbionts. In addition, the authors of Funkhouser-Jones et al. (2018) have identified specific host genes that are directly involved in the regulation of symbiont densities, thereby adding further intricacy to this association. The involvement of cellular machinery, particularly the autophagy mechanism, has been proposed as a significant factor in regulating these interactions, indicating a potentially fruitful avenue for further investigation (Cassone et al., 2015; Wang et al., 2022).

The pea aphid *Acyrthosiphon pisum* (Hemiptera: Aphididae) is a major economic pest in agriculture and forestry worldwide, known for consuming a wide range of legume plants. The pea aphid is a typical insect with alternating generations, which has the characteristics of rapid parthenogenic reproduction and a short life cycle. It not only causes discoloration, curl and deformity of plant leaves by feeding the phloem sap of host plants but also affects the normal growth and development of plants, resulting in a decline in yield and quality (Losey and Eubanks, 2000; Ryalls et al, 2013). It has the ability to transmit more than 30 plant viruses, resulting in substantial agricultural damage (Golawska and Lukasik, 2012). It is worth noting that pea aphids feed on the phloem sap of host plants, and there are limited amounts of essential amino acids and vitamins in the plant sap (Karley et al, 2002; Douglas, 2006). The core component of their biological makeup is the principal endosymbiont, *Buchnera aphidicola*. The symbiotic relationship between the aphid and its symbiont is characterized by providing a diverse range of vital nutrients that enhance the aphid's dietary requirements. The initial investigations have demonstrated the negative impact on the reproductive capacity and development of aphids when there are disturbances in this mutually beneficial association (Lv et al., 2018). Nevertheless, the intricate relationships among these symbiotic organisms, herbivorous

insects, and their host plants remain poorly understood. The current study aims to thoroughly examine the symbiotic interactions between pea aphids and *Buchnera*, focusing on their dynamics under varying host plant conditions. By comprehensively evaluating factors such as aphid weight, fecundity, *Buchnera* densities, and nutrient content, we aim to understand the influence of different host plants on the development of pea aphids and the inherent complexity of these interrelationships and to understand the different phenotypes of aphids that may be produced under the influence of different host plants. This comprehension can provide a theoretical basis for further studies on the effects of different host plants on the development of pea aphids and other insects, and provide strong support for clarifying the population evolution and species formation of aphids, and possesses the potential to illuminate the intricacies of symbiotic associations and their wider ramifications for the dynamics of ecosystems.

2 Materials and methods

2.1 Insects

The green color morphs of the pea aphid, *A. pisum*, were separately collected from alfalfa fields in Lanzhou City, Gansu Province, China. A single asexual line was established for further laboratory experiments by parthenogenesis of one pea aphid. The main host plant utilized for rearing pea aphids in our study was broad bean (*Vicia faba*). To establish the experimental population of pea aphids, pea aphids were maintained in the laboratory at a controlled temperature of 22 ± 1 °C and relative humidity of 70-80%, with a photoperiod of 16 hours of light and 8 hours of darkness (16:8 L:D). The aphids were reared on the broad bean plants for a minimum of three generations before being used in the experiments.

2.2 Host plant

In the current study, a total of six different host plant species were utilized for the experimental procedures. Expanding upon the cultivation methodologies outlined by Lv et al. (2018), the seedlings were carefully nurtured in a controlled environment. The broad bean (*V. faba*) was selected as the primary host for this research, as it provided a suitable environment for breeding our experimental pea aphid population. Furthermore, distinct populations of pea aphids were created with in a controlled laboratory setting by utilizing five different plant species, namely pea (*Pisum sativum*), alfalfa (*Medicago sativa*), clover (*Trifolium pratense*), red bean grass (*Onobrychis viciaefolia*), and melilotus (*Melilotus officinalis*). The aphid colonies were cultivated on their designated host plants for a minimum of three generations prior to their inclusion in our experiments. The host plants exhibited robust growth within a controlled greenhouse environment, where precise conditions were upheld. These conditions included a temperature range of 22 ± 1 °C, relative humidity levels ranging from 70% to 80%, and a light-dark cycle of 16 hours of light followed by 8 hours of darkness. The plants that were selected for further laboratory experiments were only those that demonstrated the highest level of resilience and durability.

2.3 Biological assay of pea aphids' weight and fecundity

The present study utilized bioassays to evaluate the influence of different host plants on the weight and reproductive capacity of pea aphids. Aphids are raised by the detached leaves-feeding method; the experiment was initiated by preparing a Petri dish with a diameter of 10 cm and then lining it with filter paper. Subsequently, a leaf in a pristine condition, positioned with its blade oriented in an upward direction, was meticulously inserted into the dish. The petioles were carefully wrapped with absorbent cotton balls that had been dampened with distilled water (ddH_2O) to ensure sufficient moisture for both the cotton balls and the filter paper. Following this, a recently hatched aphid, with age not exceeding 6 hours, was introduced into the Petri dish, enabling it to consume the leaf obtained from one of the previously mentioned six plant species. The aphids were placed individually in Petri dishes and placed in a controlled environment chamber (RZX, Ningbo Jiangnan Co. Ltd., Ningbo, China). The chamber was set to maintain a temperature of 22 ± 1 °C, a relative humidity of 70-80%, and a photoperiod of 16:8 (L:D). The aged leaves were replaced with newly harvested counterparts on a tri-daily basis. A sample size of 60 aphids was utilized for each host plant. Systematic monitoring of aphid populations on various host plants was conducted. The researchers meticulously recorded data about mortality and molting patterns, including the frequency and timing of molting, at regular 12-hour intervals. To facilitate accurate observations, the molting dander is carefully selected and collected with a camel brush to ensure precise data collection without disturbing the aphids or their environment. When the aphids developed into an adult (after the fourth molting), ten pea aphids from each host plant were randomly selected and weighed using a high-precision (1/100,000) electronic balance. Five individuals were randomly selected from each host plant to quantify *Buchnera* in the pea aphid. In addition, five pea aphids were randomly selected from the same host plant for quantitative analysis of nutrients, including total lipid, total protein, soluble sugar, and glycogen. These experiments were repeated three times. Thirty aphids were randomly selected from the same host plant to determine their reproductive capacity. Observe and record the number of nymphs produced by each aphid on a daily basis until the adult aphids died. According to the observed data, the fecundity of pea aphid on each of the six host plants was calculated.

2.4 Quantitative detection of *Buchnera*

The quantification of *Buchnera* presence in pea aphids across different host plants was conducted through the utilization of quantitative real-time PCR (q-PCR). Deoxyribonucleic acid extraction was performed on clusters of five recently matured

adult aphids for each sample in this study. The Direct-zol RNA Kits (Zymo Research, Irvine, CA, USA) were used for the isolation process. In previous studies conducted by Zhang et al. (2016) and Chang et al. (2022), the *Buchnera* 16S rRNA gene was utilized as the detection target, while the ef1α gene of the aphid was selected as the reference gene. This choice of reference gene enabled the normalization and quantification of data, as described in the studies as mentioned earlier. Detailed information on primer sequences is included in Supplementary File S1. The quantitative polymerase chain reaction (qPCR) was performed using SYBR premix Ex Taq (TaKaRa, Japan) and conducted on an ABI 7500 Real-Time PCR Detection System (Applied Biosystems/Life Technologies, USA). The qPCR mixture consisted of 5 μL of SYBR Green PCR Mastermix (TaKaRa), 1 μL of the DNA extract, 1 μL (10 μM) of the forward primer, 1 μL (10 μM) of the reverse primer, and 3 μL of sterile distilled water, resulting in a total volume of 10 μL. The thermocycling protocol consisted of an initial denaturation step at a temperature of 95°C for a duration of 30 seconds, followed by 40 cycles of denaturation at 95°C for 5 seconds, and subsequent annealing/extension at a temperature of 60°C for a duration of 34 seconds. Subsequently, a melt curve analysis was conducted, commencing at a temperature of 95°C for a duration of 15 seconds, followed by a decrease to 60°C for a period of 1 minute, and ultimately concluding at a temperature of 95°C. The experimental protocol was replicated four times.

2.5 Samples preparation for nutrient determination

A total of 90 adult aphids were collected from six different host plants to evaluate their nutritional composition, including components such as total protein, soluble sugar, glycogen, and total lipid. In order to provide further clarification regarding the methodology, a total of five mature aphids were selected from each host plant. These aphids were subjected to a thorough rinse using ddH_2O before being carefully transferred into a 1.5 mL centrifuge tube containing 180 μL of an aqueous lysis buffer formulation. The buffer solution consisted of a concentration of 100 millimolar (mM) potassium dihydrogen phosphate (KH_2PO_4), 1 mM ethylenediaminetetraacetic acid (EDTA), and 1 mM dithiothreitol (DTT). The pH of the buffer was maintained at 7.4. Subsequently, a comprehensive standardization of the aphids was performed. The resultant mixture was subjected to centrifugation at a velocity of 16,000 revolutions per minute (rpm), sustained for a duration of 15 minutes at a temperature of 4 degrees Celsius. The liquid portion obtained from this procedure was subsequently designated for subsequent nutritional evaluations. The entire procedure was repeated three times for each host plant, and the technique was replicated twice for each biological sample.

2.6 Protein content detection

The protein concentration in the pea aphid was determined using the methodologies described by Li et al. (2021) and Lv et al. (2018). Approximately 20 μL of the homogenized sample from the aphid was carefully transferred into a 1.5 mL centrifuge tube using a pipette. Subsequently, a mixture of 200 μL of coomassie brilliant blue G-250 was combined with the aforementioned sample and left to incubate for a duration of 15 minutes at room temperature. After the completion of the incubation period, the resultant mixture was introduced into a 96-well borosilicate microplate. Subsequently, spectrophotometric measurements were conducted at a wavelength of 595 nm in order to assess the protein concentration. To establish a basis for these measurements, bovine serum albumin (BSA) was dissolved in a buffer of identical composition and subsequently diluted in a stepwise manner to generate a range of concentrations. This diluted BSA solution served as a calibration standard for the experiment. The total protein content of the aphid samples was determined by employing the established curve derived from bovine serum albumin obtained from Sangon Biotech in Shanghai, China. The investigative procedure was repeated six times.

2.7 Sample preparation for the determination of soluble sugar, glycogen content, and total lipid

A volume of 180 μL of the homogenate supernatant was transferred to a new centrifuge tube with a capacity of 2 mL. A 20 μL aliquot of a 20% sodium sulfate solution was introduced to facilitate the dissolution of total carbohydrates. In order to enhance the solubility of both the total lipid and water-soluble carbohydrates, the solution was supplemented with 1,500 μL of a chloroform-methanol mixture in a 1:2 volume ratio. The composite sample was centrifugated at a speed of 10,000 revolutions per minute for 15 minutes at a temperature of 4°C. After centrifugation, the translucent liquid above the sediment was meticulously transferred using a pipette into a separate centrifuge tube. This particular tube was designated for future analysis of the overall lipid content and soluble sugar concentration. In contrast, the sediment that remained at the bottom of the tube was set aside to analyze its glycogen content.

2.8 Total lipids content detection

The methodology for determining the total lipid content in pea aphid followed procedures delineated by Handel (1965) and Lv et al. (2018). A volume of 100 μL of the previously mentioned supernatant was transferred into a centrifuge tube with a capacity of 1 mL using a pipette. The provided specimen was subsequently exposed to a temperature of 90 °C until the solvent was completely evaporated. Following that, a volume of 10 μL of concentrated sulfuric acid (98%) was added to the tube. After the inclusion of this component, the tube was subjected to incubation at a temperature of 90 °C for a short duration of 2 minutes, and subsequently rapidly cooled using ice. A 190 μL aliquot of a vanillin solution, which was prepared at a concentration of 1.2 g/L using 68% orthophosphoric acid as the solvent, was introduced into the cooled tube. The reaction was allowed to proceed for a period of 15 minutes at

room temperature. The resulting mixture was subsequently distributed into a 96-well borosilicate microplate. Spectrophotometric measurements were conducted to assess the optical density (OD) values at a specific wavelength of 525 nm. The overall lipid content was determined by utilizing the standard curve derived from triolein. The experiment was systematically replicated on six occasions.

2.9 Soluble sugar content detection

The soluble sugar concentration in pea aphids was quantified according to the methodologies described by Handel (1985) and Handel and Day (1988). In summary, a volume of approximately 150 μL of the supernatant obtained from each aphid specimen was transferred into a 1.5 mL centrifuge tube and subjected to complete evaporation under normal atmospheric conditions. Following that, a volume of 10 μL of distilled water was mixed with 240 μL of an anthrone reagent solution (with a concentration of 1.42 g/L) that had been prepared using 70% sulphuric acid. The mixture contained in the tube was allowed to incubate at ambient temperature for a duration of 15 minutes, subsequently followed by a 15-minute immersion in a boiling water bath. Following the post-heating process, the tubes were subsequently brought back to ambient temperature to facilitate the cooling process. Subsequently, the resulting mixture was transferred to a microplate made of borosilicate glass with 96 wells. Spectrophotometric measurements were conducted to obtain optical density (OD) readings at a wavelength of 630 nm. Soluble sugar concentrations were determined using a reference standard curve prepared with D-glucose. The aforementioned analytical procedure was performed on six separate occasions consistently.

2.10 Glycogen content detection

The glycogen content of the pea aphid was detected according to the method of Lv et al. (2018) and Li et al. (2021). The precipitant from pre-preparation was mixed with 400 μL of 80% methanol and made turbid by sonicating in an ultrasonic cleaning for 10 min, after which the homogenate was centrifuged again at 10,000 rpm for 5 min at 4 °C, and the supernatant was removed with a 1.5 mL centrifuge tube. Next, 1,000 μL of 1.42 g/L anthrone reagent (solvent was sulphuric acid 70%) were added to the precipitant, and the tube was incubated for 15 min at room temperature and then incubated in boiling water for a further 15 min, followed by cooling on ice to stop the reaction. After this mixture was transferred into a 96-well borosilicate microplate, the absorbance value was measured spectrophotometrically at 630 nm wavelength. As the standard, D-glucose was used to calculate the soluble sugar content. The experiment was repeated six times. The different nutrients (fats, proteins, soluble sugar, and glycogen) have been converted to energy equivalents. These reserves have energy equivalents of 39,500 mJ/mg for lipids, 24,000 mJ/mg for proteins, and 17,500 mJ/mg for soluble sugar and glycogen.

2.11 Statistical analyses

The data were organized and sorted utilizing Microsoft Excel version 2021. GraphPad Prism 8, a software provided by Systat Software Inc. in San Jose, CA, USA, was utilized to create subsequent graphical representations. In order to conduct statistical analyses, the researchers utilized IBM SPSS Statistics version 22.0 (IBM, Armonk, NY, USA). In order to identify noteworthy differences among different treatments, the statistical technique of analysis of variance (ANOVA) was employed, specifically utilizing Tukey's honestly significant difference (HSD) test.

3 Results

3.1 Effects of different host plants on the weight of pea aphid

The six host plants affect the weight of the pea aphid differently. Compared with other host plants, the weight of a pea aphid fed on *V. faba* was the highest, and the weight of a pea aphid fed on *M. sativa* was the lowest. However, there were no significant differences in weight that fed on *M. officinalis*, *T. pratense*, *P. sativum*, and *O. viciaefolia* (Figure 1, F (5, 174) = 43.185, $P < 0.001$). This indicates that pea aphid has good adaptability on *V. faba*.

3.2 The fecundity of pea aphids on different host plants

Pea aphids have different fecundity on different host plants; the pea aphid had the highest fecundity fed on *V. faba* in six host plants, which was a significant difference compared to other host plants. The fecundity of pea aphids on *O. viciaefolia* was the lowest in six host plants, but pea aphis showed no significant difference from *M. sativa*. The fecundity of pea aphids on *M. officinalis* significantly differed from the other host plants. In addition, there was no significant difference in the fecundity of pea aphids between *P. sativum* and *T. pratense*. The data showed that among pea aphids, feeding on *V. faba* is beneficial to the fecundity of aphids. (Figure 2, F (5, 174) = 43.185, $P < 0.001$).

3.3 Titer mensuration of *Buchnera* in the pea aphid

The relative expression of *Buchnera* in the pea aphid was variable depending on the host plant. The relative expression of *Buchnera* in pea aphids fed on *T. pratense* was significantly higher than that of *V. faba*, *P. sativum*, *M. officinalis*, and *O. viciaefolia*. The relative expression of *Buchnera* in pea aphids was significantly different in the *V. faba* and *O. viciaefolia* compared with other host plants, but not significantly from each other (Figure 3, F (5, 18) = 29.873, $P < 0.001$). The results showed that host plants significantly affected the titer of symbiotic bacteria in pea aphids, which may be related to the nutritional status of host plants.

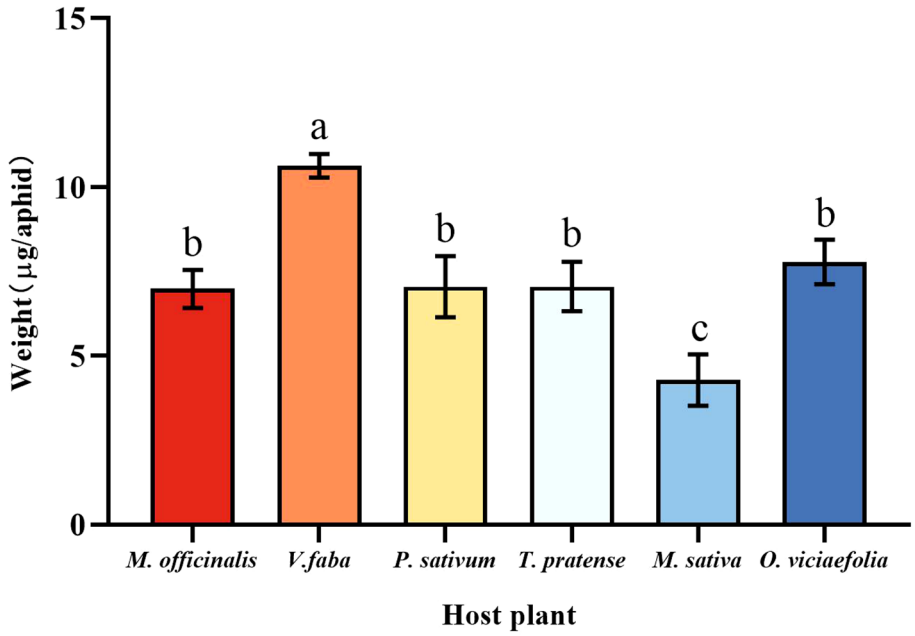


FIGURE 1
Effects of different host plants on the weight of pea aphids. Different lowercase letters between different host plants are significant differences as determined by one-way ANOVA Tukey's HSD test at $p < 0.05$.

3.4 The protein content of different host plants of pea aphids

The total protein content of pea aphids was different fed on the different host plants. The total protein content of pea aphids was significantly different in the *M. officinalis* and *V. faba* compared with other host plants, but not significantly different from each other. However, the total protein content of pea aphid was not significantly different among the *T. pratense*, *P. sativum*, *O. viciaefolia*, and *M. sativa*. (Figure 4, F (5, 30) = 8.611, $P < 0.001$). It can be seen that feeding on *M. officinalis* and *V. faba* improves the protein content in pea aphids.

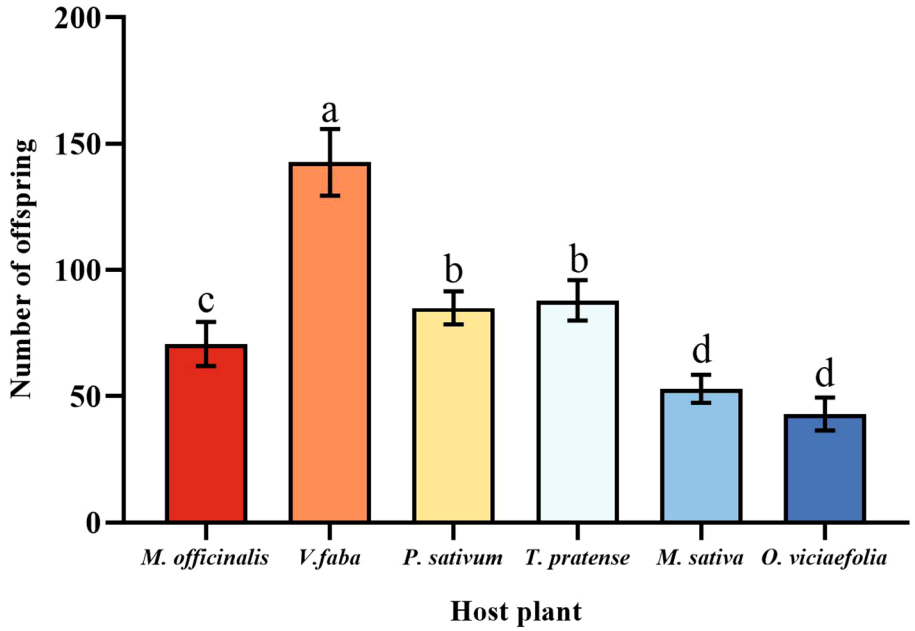


FIGURE 2
Effects of different host plants on the fecundity of pea aphids. Different lowercase letters between different host plants are significant differences as determined by one-way ANOVA Tukey's HSD test at $p < 0.05$.

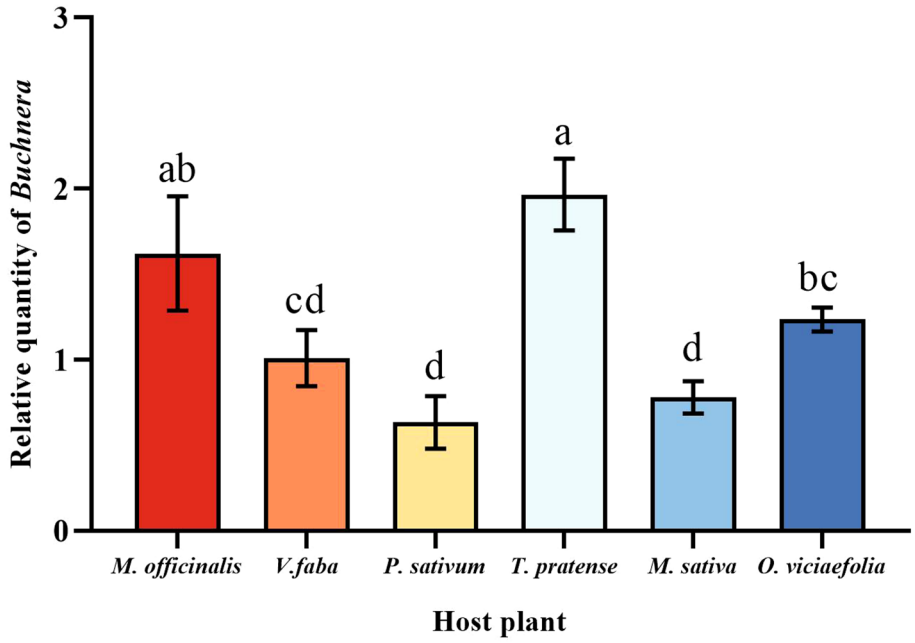


FIGURE 3
Effects of different host plants on the *Buchnera* titer in pea aphids. Different lowercase letters between different host plants are significant differences as determined by one-way ANOVA Tukey's HSD test at $p < 0.05$.

3.5 Quantification of different host plants of pea aphids to soluble sugar

The soluble sugar level of pea aphids varied among the six host plants. The maximum quantity of soluble sugar of pea aphid was fed on *V. faba*, which differed statistically from other plants. The soluble sugar content of the pea aphid fed on *T. pratense* was the lowest, which is not statistically different from *M. officinalis*. Moreover, the soluble sugar content of the pea aphids among *P. sativum*, *M. officinalis*, *M. sativa*, and *O. viciaefolia* showed no significant differences, which indicated that the soluble sugar

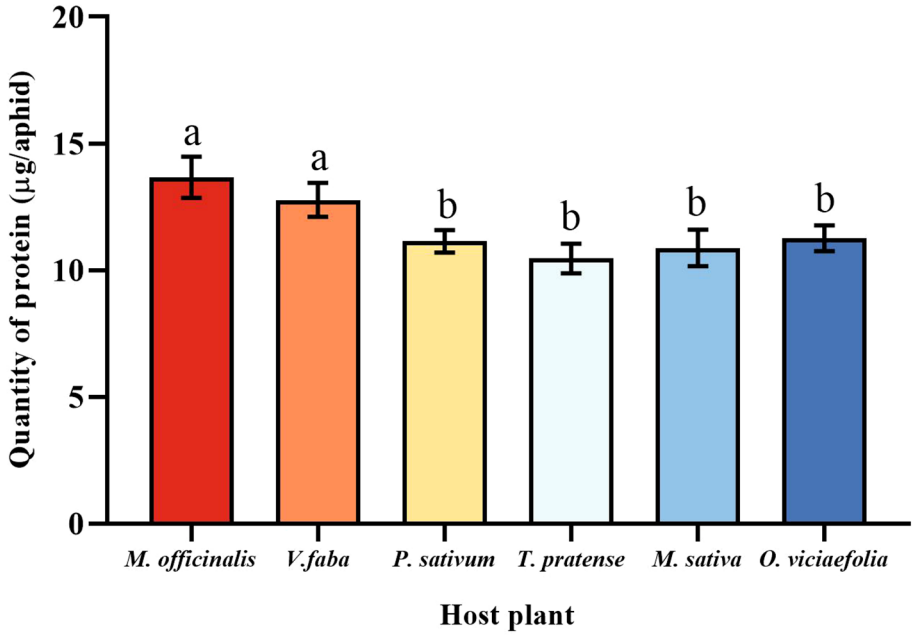


FIGURE 4
Effects of different host plants on the protein quantity of pea aphid. Different lowercase letters between different host plants are significant differences as determined by one-way ANOVA Tukey's HSD test at $p < 0.05$.

content in the pea aphid is higher after feeding on *V. faba*. (Figure 5, F (5, 30) = 16.896, $P < 0.001$).

3.6 Glycogen content of different host plants of the pea aphid

Glycogen and soluble sugar are important carbohydrates in aphids, the different host plants of the pea aphid have different patterns in the content of the two carbohydrates. The glycogen of pea aphids fed on *V. faba* and *O. viciaefolia* was significantly higher than that of the other host plants, but the differences between them were insignificant. There were no significant differences in glycogen of pea aphids among *P. sativum*, *M. officinalis*, *M. sativa*, and *T. pratense* had no significant differences (Figure 6, F (5, 30) = 6.233, $P < 0.001$), suggesting that feeding on *V. faba* and *O. viciaefolia* was conducive to the increase of glycogen in pea aphids.

3.7 Quantification of different host plants of pea aphids to total lipids

The total lipids content of the pea aphid was variable depending on the host plant. The total lipids content of the pea aphid fed on *T. pratense* and *P. sativum* were significantly greater than that of the *V. faba*, *M. officinalis*, and *M. sativa*. The total lipids content of pea aphid was significantly different in the *M. officinalis* and *V. faba* compared with other host plants, but not significantly different from each other (Figure 7, F (5, 30) = 10.545, $P < 0.001$).

3.8 Effects of different host plants on the total energy of pea aphid

The maximum total energy of pea aphid was fed on *V. faba*, which differed statistically from other plants. The total energy of the pea aphid fed on *T. pratense* was the lowest, which is not statistically different from *M. officinalis*. Furthermore, the total energy of the pea aphids among *P. sativum*, *M. officinalis*, *M. sativa*, and *O. viciaefolia* showed no significant differences (Figure 8, F (5, 30) = 16.328, $P < 0.001$).

3.9 Correlation between *Buchnera* titer and biological characteristics and nutrient metabolism of the pea aphid

Correlation analysis found that the *Buchnera* titer of pea aphids fed on different host plants had different correlations with biological characteristics and nutritional metabolism (Table 1). The difference in *Buchnera* titer was positively correlated with the protein content in *M. officinalis* and total energy content in *T. pratense* ($P < 0.05$). On the other four host plants, the titer of *Buchnera* was correlated to biological characteristics and nutritional metabolism, but it has not reached a significant level. It suggests that the host plant influences the biological characteristics, nutrient metabolism and content of symbiotic bacteria in the pea aphid and that the titer of symbiotic bacteria in the pea aphid is actively regulated according to the nutrient status of the host plant.

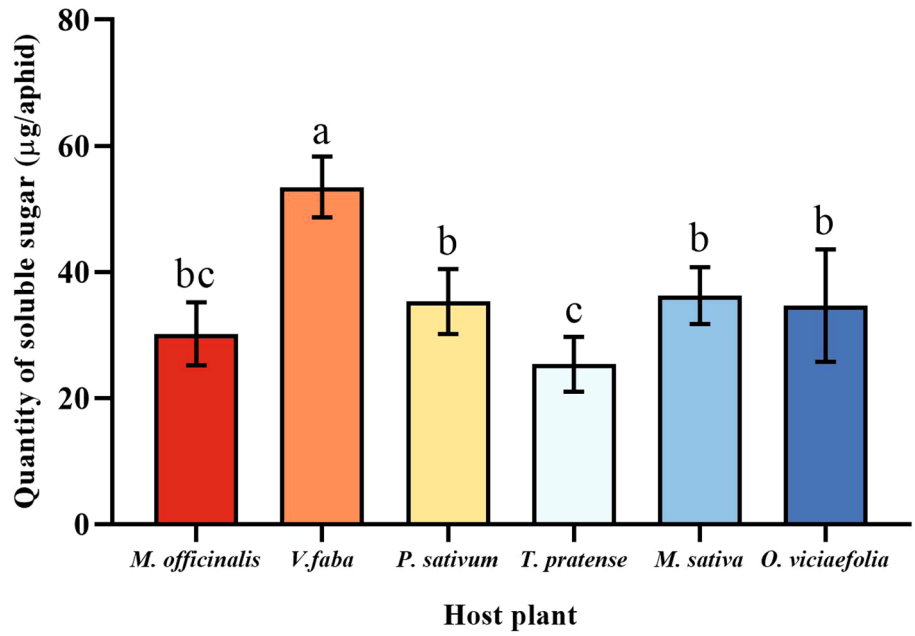


FIGURE 5
Effects of different host plants on the soluble sugar quantity of pea aphids. Different lowercase letters between different host plants are significant differences as determined by one-way ANOVA Tukey's HSD test at $p < 0.05$.

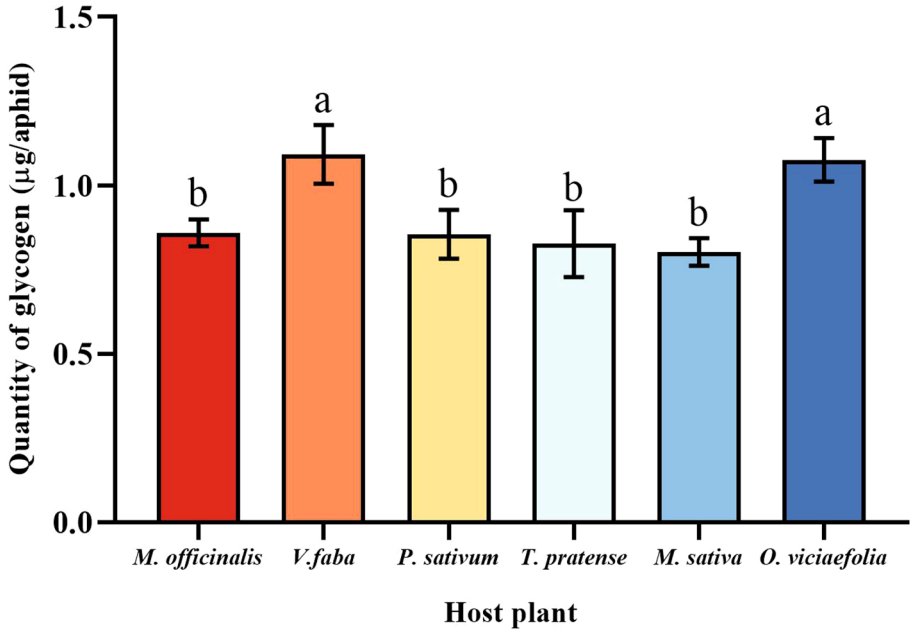


FIGURE 6
Effects of different host plants on the glycogen quantity of pea aphids. Different lowercase letters between different host plants are significant differences as determined by one-way ANOVA Tukey's HSD test at $p < 0.05$.

4 Discussion

Aphids use mouthparts to obtain nutrients from the phloem sieve elements of plants, but due to the absence of essential amino acids, the main source of nutrients does not satisfy the development and reproduction of aphids (Nalam et al., 2019). Symbiotic bacteria synthesize vitamins and other nutrients for the host insect, which affects the growth and development of the host insect. (Russell et al., 2014; Manzano-Marı et al., 2020). This study found that pea aphids differed in body weight and fecundity after feeding on different host plants, showing higher fecundity and weight gain on *V. faba*. This is

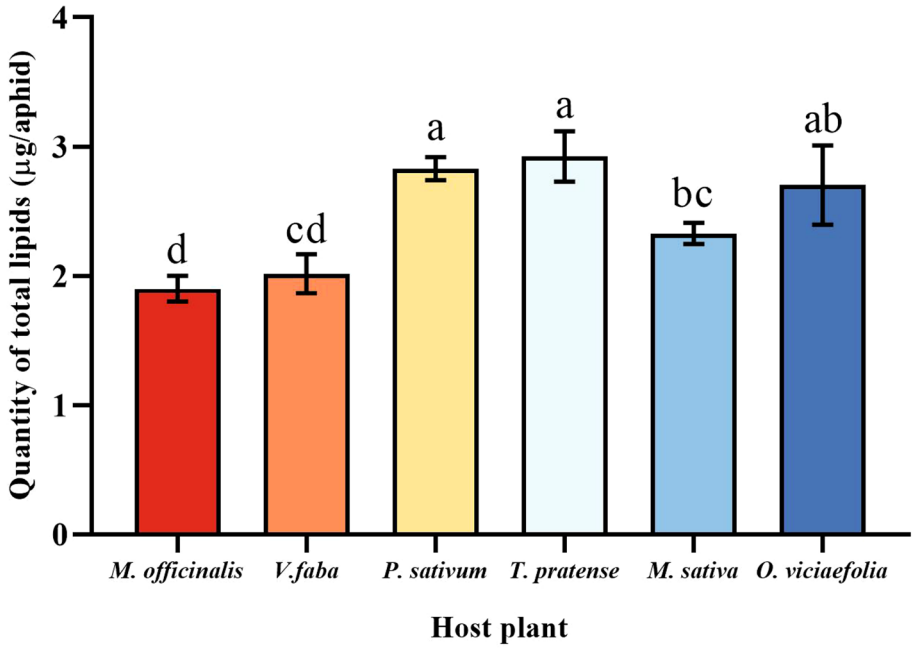


FIGURE 7
Effects of different host plants on the total lipid quantity of pea aphid. Different lowercase letters between different host plants are significant differences as determined by one-way ANOVA Tukey's HSD test at $p < 0.05$.

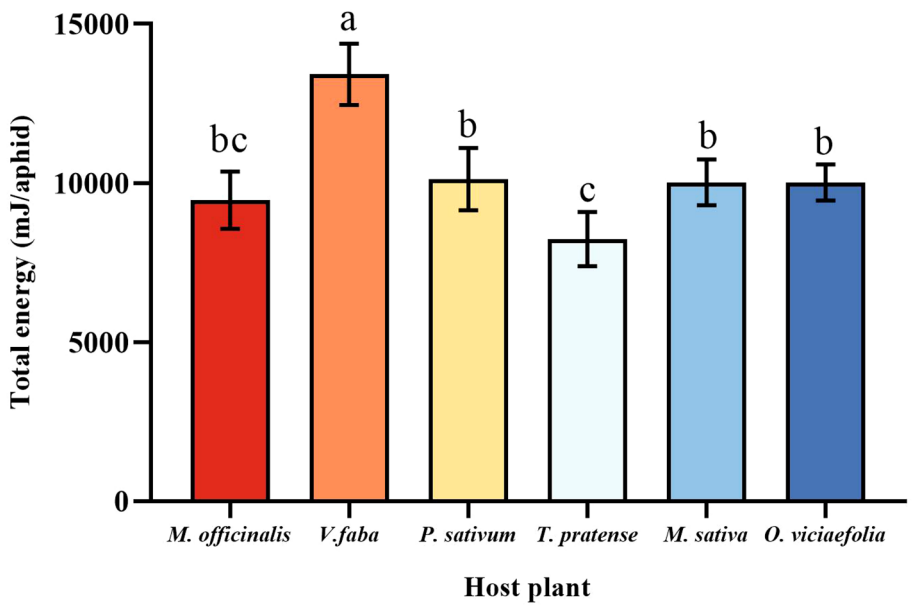


FIGURE 8
Effects of different host plants on the total energy of pea aphid. Different lowercase letters between different host plants are significant differences as determined by one-way ANOVA Tukey's HSD test at $p < 0.05$.

similar to the findings of Han et al. (2012). This may be related to the differences in insect development, morphology and physiological characteristics caused by different host plants (Peccoud et al., 2010). This shows that feeding on *V. faba* is conducive to the growth and development of pea aphids. This also shows that different host plants have different effects on the growth of aphids, and *Buchnera* is very important for aphids to obtain nutrients on the host (Xu et al., 2023).

Symbiotic bacteria not only obtain and recover nitrogen-containing precursors from the food and nitrogenous metabolic substances of the host (Wang et al., 2020). It can also provide substances such as pectinase to help the host feed, facilitating the absorption and metabolism of nutrients by the host (Salem et al., 2017). Previous studies have found that the elimination of the *Buchnera* will reduce the protein contents of pea aphids, and the fecundity of pea aphids is in the direct ratio to its protein content: the higher the protein content, the higher the fecundity (Lv et al., 2018). So, the titer of the *Buchnera* was inextricably linked with the biological characteristics of the pea aphid. However, this experiment revealed found that the *Buchnera* titers of pea aphid on different hosts was different, among which the *Buchnera* titer on *T. pratense* was the highest. Thairu and Hansen (2019) that *Buchnera* sRNAs were differentially expressed when aphids fed on different plants. There are three reasons for this: firstly, Phytotoxins or secondary metabolites produced by different host plants can inhibit or promote the increase of the symbiotic bacteria titer in pea

TABLE 1 Correlation between the *Buchnera* titer and the biological characteristics and nutritional metabolism of pea aphids on different host plants.

Index	*M. officinalis*		*V. faba*		*P. sativum*		*T. pratense*		*M. sativa*		*O. viciaefolia*	
	Correlation coefficient	*P*	Correlation coefficient	*P*	Correlation coefficient	*P*	Correlation coefficient	*P*	Correlation coefficient	*P*	Correlation coefficient	*P*
Weight	-0.994	0.070	0.874	0.323	0.712	0.496	-0.698	0.508	0.966	0.167	-0.992	0.080
Fecundity	-0.989	0.095	-0.646	0.553	0.981	0.123	-0.733	0.476	-0.863	0.174	0.296	0.808
Protein	0.978*	0.022	-0.816	0.184	-0.900	0.100	0.033	0.967	0.270	0.730	-0.253	0.747
Soluble sugar	0.777	0.223	-0.325	0.675	-0.893	0.107	0.424	0.576	-0.913	0.087	0.778	0.222
Total lipids	-0.779	0.221	-0.285	0.715	-0.151	0.849	0.535	0.465	-0.267	0.733	-0.898	0.102
Glycogen	-0.669	0.331	0.366	0.634	0.853	0.147	0.561	0.439	0.075	0.925	-0.644	0.356
Total energy	0.892	0.108	-0.361	0.639	-0.883	0.117	0.981*	0.019	-0.949	0.051	0.753	0.247

* indicates significant relevancy at 0.05 level.

aphids, and secondary metabolites also have an inhibitory and bactericidal effect (Pawar et al., 2022). Secondly, significant differences in the amino acid composition of different host plants can interfere with the optimal symbiont density (Sandstrom and Pettersson, 1994; Douglas, 2003). Thirdly, aphids actively adjust their symbiotic density according to the nutritional status of the different host plants (Guilhot et al., 2020). Therefore, feeding on different host plants will directly affect the growth, development and nutritional metabolism of pea aphids. The results further verified the hypothesis that host plants will affect the symbiotic bacteria of phytophagous insects (Zhang et al., 2016).

Protein, carbohydrates, glycogen, and lipids, as the main forms of insects' energy storage, have an important influence on the development and life activities of insects (Ahsaei et al., 2014). Saccharides and lipids are used as a source of energy for movement, providing energy for the host's muscles as the insect walks or escapes (Hansford and Johnson, 1975; Martin and Lieb, 1979). Previous studies have demonstrated that eliminating *Buchnera* in pea aphid not only reduces the content of protein, but also slows down the feeding behavior of the pea aphid (Lv et al., 2018). This study found that differences in *Buchnera* levels in the pea aphid were due to feeding on different host plants, which affected the levels of saccharides and the insect's behavior. It is indicated that *Buchnera* can also affect the feeding of pea aphids and their adaptability to host plants by participating in the interaction between plants and aphids (MaChado-Assefh et al., 2015; MaChado-Assefh and Alvarez, 2018). In addition, the total lipid and energy content of pea aphid are also different on different hosts, which may be because the saccharides in the aphid eventually store energy in the form of lipids (Febvay et al., 1999). Therefore, feeding on different hosts can lead to a difference in the *Buchnera* titer, showing the difference in the nutritional metabolism of pea aphids. However, in many insect systems, the density of obligate symbiosis is actively adjusted by the host to cope with the changes in environmental and physiological factors (Whittle et al., 2021). This experiment found that the correlation between the titer of *Buchnera* and the life activities and nutritional metabolism of pea aphids varied with host plants. This shows that insect hosts will actively adjust the symbiotic density according to the availability of nutrients and dietary requirements (Wilkinson et al., 2007; Snyder et al., 2012). Therefore, insect hosts benefit from symbiosis with obligate symbionts and have costs related to providing energy and nutrients to maintain the symbiotic population (Engl et al., 2020).

To summarize, host plants not only affect the biological characteristics and nutritional metabolism of pea aphids, but also regulated the symbiotic density. A previous study found that the density of *Buchnera* in cotton aphids differed among aphid populations fed on different host plants (Zhang et al., 2016). However, the titer of symbiotic bacteria is closely related to the nutritional metabolism of pea aphids, so different host plants will further influence the biological characteristics and nutritional metabolism of pea aphids by affecting the titer of symbiotic bacteria. This study demonstrates that among the six host plants, pea aphid has good adaptability on *V. faba*, and different host plants will also interfere with the nutritional function of *Buchnera*. These changes may lead to the transfer of energy level of pea aphids, thus affecting the growth and development of pea aphids. However, the changes in nutritional requirements of pea aphids fed on different host plants and the cost of symbiosis with obligate symbionts need further study. In addition, aphids with different phenotypes may be produced, accelerating the population evolution and species formation of aphids and responding to the changes in the external environment. Because of the geographical latitude difference, temperature and light work together in nature. Therefore, on the basis of this study, field experiments were carried out to explore the effects of the external natural environment on the growth, development and nutritional metabolism of aphids with different phenotypes and clarify the symbiotic-mediated population evolution and speciation of herbivorous insects. It provides a theoretical basis for further study on population evolution and species formation of pea aphid, and provides powerful information for designing better control strategies.

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

H-PL: Investigation, Formal analysis, Validation, Writing – original draft. Q-YY: Investigation, Data curation, Formal analysis, Writing – original draft. J-XL: Resources, Validation, Data curation. IH: Data curation, Writing – review & editing. YL: Formal analysis, Resources. Q-YZ: Resources, Data curation. KA: Writing – review & editing. AA: Writing – review & editing. C-ZL: Conceptualization, Resources, Supervision. NL: Conceptualization, Funding acquisition, Supervision, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was funded by the National Natural Science Foundation of China (No. 32060250).

Acknowledgments

The authors extend their appreciation to the Researchers Supporting Project number (RSP-2024R369), King Saud University, Riyadh, Saudi Arabia. Many thanks to the Insect Ecology Laboratory Ph.D. Ke-xin Zhang (Gansu Agricultural University) for his technology help and English language revision during the preparation of this manuscript.

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1288997/full#supplementary-material

References

Ahsaei, S. M., Hosseininaveh, V., Talaei-Hassanlouei, R., and Bigham, M. (2014). Changes in energy content of *Podisus maculiventris* (Say) (Hemiptera: Pentatomidae) in response to different diets. *Arthropods.* 34, 166–173.

Bao, X. Y., Yan, J. Y., Yao, Y. L., Wang, Y. B., Visendi, P., Seal, S., et al. (2021). Lysine provisioning by horizontally acquired genes promotes mutual dependence between whitefly and two intracellular symbionts. *PloS Pathog.* 17, e1010120. doi: 10.1371/journal.ppat.1010120

Baumann, P. (2005). Biology bacteriocyte-associated endosymbionts of plant sap-sucking insects. *Annu. Rev. Microbiol.* 59, 155–189. doi: 10.1146/annurev.micro.59.030804.121041

Bulman, C. A., Chappell, L., Gunderson, E., Vogel, I., Beerntsen, B., Slatko, B. E., et al. (2021). The Eagle effect in the *Wolbachia*-worm symbiosis. *Parasit Vectors.* 14, 118. doi: 10.1186/s13071-020-04545-w

Cassone, B. J., Redinbaugh, M. G., Dorrance, A. E., and Michel, A. P. (2015). Shifts in *Buchnera aphidicola* density in soybean aphids (*Aphis glycines*) feeding on virus-infected soybean. *Insect Mol. Biol.* 24, 422–431. doi: 10.1111/imb.12170

Chang, C. Y., Sun, X. W., Tian, P. P., Miao, N. H., Zhang, L. Y., and Liu, X. D. (2022). Plant secondary metabolite and temperature determine the prevalence of *Arsenophonus* endosymbionts in aphid populations. *Environ. Microbiol.* 24, 3764–3776. doi: 10.1111/1462-2920.15929

Deehan, M., Lin, W., Blum, B., Emili, A., and Frydman, H. (2021). Intracellular density of *Wolbachia* is mediated by host autophagy and the bacterial cytoplasmic incompatibility gene *cifB* in a cell type-dependent manner in *Drosophila melanogaster*. *mBio.* 12, e02205–e02220. doi: 10.1128/mBio.02205-20

Douglas, A. E. (1998). Nutritional interactions in insect-microbial symbioses: aphids and their symbiotic bacteria *Buchnera*. *Annu. Rev. Entomol.* 43, 17–37. doi: 10.1146/annurev.ento.43.1.17

Douglas, A. E. (2003). The nutritional physiology of aphids. *Adv. Insect Physiol.* 31, 73–140. doi: 10.1016/s0065-2806(03)31002-1

Douglas, A. E. (2006). Phloem-sap feeding by animals: Problems and solutions. *J. Exp. Bot.* 57, 747–754. doi: 10.1093/jxb/erj067

Douglas, A. E. (2015). Multiorganismal insects: diversity and function of resident microorganisms. *Annu. Rev. Entomol.* 60, 17–34. doi: 10.1146/annurev-ento-010814-020822

Duron, O., Morel, O., Noël, V., Buysse, M., Binetruy, F., Lancelot, R., et al. (2018). Tick-Bacteria mutualism depends on B vitamin synthesis pathways. *Curr. Biol.* 28, 1896–1902.e5. doi: 10.1016/j.cub.2018.04.038

Engl, T., Schmidt, T. H. P., Kanyile, S. N., and Klebsch, D. (2020). Metabolic cost of a nutritional symbiont manifests in delayed reproduction in a grain pest beetle. *Insects.* 11, 1–15. doi: 10.3390/insects11100717

Febvay, G., Rahbe, Y., Rynkiewicz, M., Guillaud, J., and Bonnot, G. (1999). Fate of dietary sucrose and neosynthesis of amino acids in the pea aphid, *Acyrthosiphon pisum*, reared on different diets. *J. Exp. Biol.* 202, 2639–2652. doi: 10.1242/jeb.202.19.2639

Funkhouser-Jones, L. J., van Opstal, E. J., Sharma, A., and Bordenstein, S. R. (2018). The maternal effect gene *Wds* controls *Wolbachia* titer in *Nasonia*. *Curr. Biol.* 28, 1692–1702.e6. doi: 10.1016/j.cub.2018.04.010

Golawska, S., and Lukasik, I. (2012). Antifeedant activity of luteolin and genistein against the pea aphid, *Acyrthosiphon pisum*. *J. Pest. Sci.* 85, 443–450. doi: 10.1007/s10340-012-0452-z

Guilhot, R., Rombaut, A., Xuereb, A., Howell, K., and Fellous, S. (2020). Environmental specificity in Drosophila-bacteria symbiosis affects host developmental plasticity. *Evol. Ecol.* 34, 693–712. doi: 10.1007/s10682-020-10068-8

Han, X. N., Wang, X. Q., Zhao, L. P., and Liu, C. Z. (2012). Effect of host plants on the growth, development and fecundity of *Acyrthosiphon pisum*. *Plant Protection.* 8, 40–43. doi: 10.3969/j.issn.0529-1542.2012.01.008

Handel, V. E. (1965). Microseparation of glycogen, sugars, and lipids. *Anal. Biochem.* 11, 266–271. doi: 10.1016/0003-2697(65)90014-x

Handel, V. E. (1985). Rapid determination of total lipids in mosquitoes. *J. Am. Mosq. Control Assoc.* 1, 302–304.

Handel, V. E., and Day, J. (1988). Assay of lipids, glycogen and sugars in individual mosquitoes: correlations with wing length in field-collected *Aedes vexans*. *J. Am. Mosq. Control Assoc.* 4, 549–550.

Hansford, R. G., and Johnson, R. N. (1975). The nature and control of the tricarboxylate cycle in beetle flight muscle. *Biochem. J.* 148, 389–401. doi: 10.1042/bj1480389

Harumoto, T., Anbutsu, H., Lemaitre, B., and Fukatsu, T. (2016). Male-killing symbiont damages host's dosage-compensated sex chromosome to induce embryonic apoptosis. *Nat. Commun.* 7, 12781. doi: 10.1038/ncomms12781

Himler, A. G., Adachi-Hagimori, T., Bergen, J. E., Kozuch, A., Kelly, S. E., Tabashnik, B. E., et al. (2011). Rapid spread of a bacterial symbiont in an invasive whitefly is driven by fitness benefits and female bias. *Science.* 332, 254–256. doi: 10.1126/science.1199410

Karley, A. J., Douglas, A. E., and Parker, W. E. (2002). Amino acid composition and nutritional quality of potato leaf phloem sap for aphids. *J. Exp. Biol.* 205, 3009–3018. doi: 10.1242/jeb.205.19.3009

Li, C. C., Sun, Q., Gou, Y. P., Zhang, K. X., Zhang, Q. Y., Zhou, J. J., et al. (2021). Long-term effect of elevated CO_2 on the development and nutrition contents of the pea aphid (*Acyrthosiphon pisum*). *Front. Physiol.* 12, 688220. doi: 10.3389/fphys.2021.688220

Losey, J. E., and Eubanks, M. D. (2000). Implications of pea aphid host-plant specialization for the potential colonization of vegetables following post-harvest emigration from forage crops. *Environ. Entomol.* 29, 1283–1288. doi: 10.1603/0046-225X-29.6.1283

Lv, N., Peng, J., Chen, X. Y., Guo, C. F., Sang, W., Wang, X. M., et al. (2021). Antagonistic interaction between male-killing and cytoplasmic incompatibility induced by *Cardinium* and *Wolbachia* in the whitefly, *Bemisia tabaci*. *Insect Sci.* 28, 330–346. doi: 10.1111/1744-7917.12793

Lv, N., Wang, L., Sang, W., Liu, C. Z., and Qiu, B. L. (2018). Effects of endosymbiont disruption on the nutritional dynamics of the pea aphid *Acyrthosiphon pisum*. *Insects.* 9, 161. doi: 10.3390/insects9040161

MaChado-Assefh, C. R., and Alvarez, A. E. (2018). Probing behavior of aposymbiotic green peach aphid (*Myzus persicae*) on susceptible Solanum tuberosum and resistant Solanum stoloniferum plants. *Insect Sci.* 25(1), 127–136. doi: 10.1111/1744-7917.12372

MaChado-Assefh, C. R., Lopez-Isasmendi, G., Tjallingii, W. F., Jander, G., and Alvarez, A. E. (2015). Disrupting *Buchnera aphidicola*, the endosymbiotic bacteria of *Myzus persicae*, delays host plant acceptance. *Arthropod Plant Inte.* 9, 529–541. doi: 10.1007/s11829-015-9394-8

Manzano-Marı, N. A., Coeur d'acier, A., Clamens, A. L., Orvain, C., Cruaud, C., Barbe, V., et al. (2020). Serial horizontal transfer of vitamin-biosynthetic genes enables the establishment of new nutritional symbionts in aphids' di-symbiotic systems. *ISME J.* 14, 259–273. doi: 10.1038/s41396-019-0533-6

Martin, M. M., and Lieb, T. J. (1979). Patterns of fuel utilization by the thoracic muscles of adult worker ants. The use of lipid by a hymenopteran. *Comp. Biochem. Physiol. Part B Comp. Biochem.* 64, 387–390. doi: 10.1016/0305-0491(79)90287-6

Moran, N. A., and Bennett, G. M. (2014). The tiniest tiny genomes. *Annu. Rev. Microbiol.* 68, 195–215. doi: 10.1146/annurev-micro-091213-112901

Moran, N. A., McCutcheon, J. P., and Nakabachi, A. (2008). Genomics and evolution of heritable bacterial symbionts. *Annu. Rev. Genet.* 42, 165–190. doi: 10.1146/annurev.genet.41.110306.130119

Morrow, J. L., Hall, A. A. G., and Riegler, M. (2017). Symbionts in waiting: the dynamics of incipient endosymbiont complementation and replacement in minimal bacterial communities of psyllids. *Microbiome.* 5, 58. doi: 10.1186/s40168-017-0276-4

Nalam, V., Louis, J., and Shah, J. (2019). Plant defense against aphids, the pest extraordinaire. *Plant Sci.* 279, 96–107. doi: 10.1016/j.plantsci.2018.04.027

Nguyen, D. T., Morrow, J. L., Spooner-Hart, R. N., and Riegler, M. (2017). Independent cytoplasmic incompatibility induced by *Cardinium* and *Wolbachia* maintains endosymbiont coinfections in haplodiploid thrips populations. *Evolution.* 71, 995–1008. doi: 10.1111/evo.13197

Patton, M. F., Hansen, A. K., and Casteel, C. L. (2021). Potato leafroll virus reduces *Buchnera aphidicola* titer and alters vector transcriptome responses. *Sci. Rep.* 11, 23931. doi: 10.1038/s41598-021-02673-6

Pawar, M. M., Shivanna, B., Prasannakumar, M. K., Parivallal, P. B., Suresh, K., and Meenakshi, N. H. (2022). Spatial distribution and community structure of microbiota associated with cowpea aphid (*Aphis craccivora* Koch). *3 Biotech.* 12, 75. doi: 10.1007/s13205-022-03142-1

Peccoud, J., Simon, J. C., von Dohlen, C., d'acier, A., Plantegenest, M., and Vanlerberghe-Masutti, F. (2010). Evolutionary history of aphid-plant associations and their role in aphid diversification. *C R Biol.* 333, 474–487. doi: 10.1016/j.crvi.2010.03.004

Perlmutter, J. I., and Bordenstein, S. R. (2020). Microorganisms in the reproductive tissues of arthropods. *Nat. Rev. Microbiol.* 18, 97–111. doi: 10.1038/s41579-019-0309-z

Russell, C. W., Poliakov, A., Haribal, M., Jander, G., van Wijk, K. J., and Douglas, A. E. (2014). Matching the supply of bacterial nutrients to the nutritional demand of the animal host. *Proc. Biol. Sci.* 281, 20141163. doi: 10.1098/rspb.2014.1163

Ryalls, J. M. W., Riegler, M., Moore, B. D., and Johnson, S. N. (2013). Biology and trophic interactions of lucerne aphids. *Agric. For. Entomol.* 15, 335–350. doi: 10.1111/afe.12024

Salem, H., Bauer, E., Kirsch, R., Berasategui, A., Cripps, M., Weiss, B., et al. (2017). Drastic genome reduction in an herbivore's pectinolytic symbiont. *Cell.* 171, 1520–1531.e13. doi: 10.1016/j.cell.2017.10.029

Sandstrom, J., and Pettersson, J. (1994). Amino acid composition of phloem sap and the relation to intraspecific variation in pea aphid (*Acyrthosiphon pisum*) performance. *J. Insect Physiol.* 40, 947–955. doi: 10.1016/0022-1910(94)90133-3

Sloan, D. B., Nakabachi, A., Richards, S., Qu, J., Murali, S. C., Gibbs, R. A., et al. (2014). Parallel histories of horizontal gene transfer facilitated extreme reduction of endosymbiont genomes in sap-feeding insects. *Mol. Biol. Evol.* 31, 857–871. doi: 10.1093/molbev/msu004

Snyder, A. K., McLain, C., and Rio, R. V. (2012). The tsetse fly obligate mutualist *Wigglesworthia* morsitans alters gene expression and population density via exogenous nutrient provisioning. *Appl. Environ. Microbiol.* 78, 7792–7797. doi: 10.1128/AEM.02052-12

Stouthamer, R., Breeuwer, J. A., and Hurst, G. D. (1999). *Wolbachia pipientis*: microbial manipulator of arthropod reproduction. *Annu. Rev. Microbiol.* 53, 71–102. doi: 10.1146/annurev.micro.53.1.71

Thairu, M. W., and Hansen, A. K. (2019). Changes in aphid host plant diet influence the small-RNA expression profiles of its obligate nutritional symbiont, *Buchnera*. *mBio.* 10, e01733–e01719. doi: 10.1128/mBio.01733-19

Vostinar, A. E., Skocelas, K. G., Lalejini, A., and Zaman, L. (2021). Symbiosis in digital evolution: past, present, and future. *Front. Ecol. Evol.* 9. doi: 10.3389/fevo.2021.739047

Wang, S., Wang, L., Fan, X., Yu, C., Feng, L., and Yi, L. (2020). An insight into diversity and functionalities of gut microbiota in insects. *Curr. Microbiol.* 77, 1976–1986. doi: 10.1007/s00284-020-02084-2

Wang, Y. B., Li, C., Yan, J. Y., Wang, T. Y., Yao, Y. L., Ren, F. R., et al. (2022). Autophagy regulates whitefly-symbiont metabolic interactions. *Appl. Environ. Microbiol.* 88, e0208921. doi: 10.1128/AEM.02089-21

Weinert, L. A., Araujo-Jnr, E. V., Ahmed, M. Z., and Welch, J. J. (2015). The incidence of bacterial endosymbionts in terrestrial arthropods. *Proc. Bio. Sci.* 282, 20150249. doi: 10.1098/rspb.2015.0249

Werren, J. H., Baldo, L., and Clark, M. E. (2008). *Wolbachia*: master manipulators of invertebrate biology. *Nat. Rev. Microbiol.* 6, 741–751. doi: 10.1038/nrmicro1969

Whittle, M., Barreaux, A. M. G., Bonsall, M. B., Ponton, F., and English, S. (2021). Insect-host control of obligate, intracellular symbiont density. *Proc. Biol. Sci.* 288, 20211993. doi: 10.1098/rspb.2021.1993

Wilkinson, T. L., Koga, R., and Fukatsu, T. (2007). Role of host nutrition in symbiont regulation: impact of dietary nitrogen on proliferation of obligate and facultative bacterial endosymbionts of the pea aphid. *Appl. Environ. Microbiol.* 73, 1362–1366. doi: 10.1128/AEM.01211-06

Wilson, A. C. C., and Duncan, R. P. (2015). Signatures of host/symbiont genome coevolution in insect nutritional endosymbioses. *Proc. Natl. Acad. Sci. U.S.A.* 112, 10255–10261. doi: 10.1073/pnas.1423305112

Xu, W., Liu, W., Li, J., Zhu, X., Wang, L., and Li, D. (2023). *Buchnera* breaks the specialization of the cotton-specialized aphid (*Aphis gossypii*) by providing nutrition through zucchini. *Front. Nutr.* 10, 1128272. doi: 10.3389/fnut.2023.1128272

Zhang, Y. C., Cao, W. J., Zhong, L. R., Godfray, H. C. J., and Liu, X. D. (2016). Host plant determines the population size of an obligate symbiont (*Buchnera aphidicola*) in aphids. *Appl. Environ. Microbiol.* 82, 2336–2346. doi: 10.1128/AEM.04131-15

Zhang, B., Leonard, S. P., Li, Y., and Moran, N. A. (2019). Obligate bacterial endosymbionts limit thermal tolerance of insect host species. *Proc. Natl. Acad. Sci. U.S.A.* 116, 24712–24718. doi: 10.1073/pnas.1915307116

Zhao, D. X., Zhang, Z. C., Niu, H. T., and Guo, H. F. (2020). Selective and stable elimination of endosymbionts from multiple-infected whitefly *Bemisia tabaci* by feeding on a cotton plant cultured in antibiotic solutions. *Insect Sci.* 27, 964–974. doi: 10.1111/1744-7917.12703

Zhu, D. T., Rao, Q., Zou, C., Ban, F. X., Zhao, J. J., and Liu, S. S. (2022). Genomic and transcriptomic analyses reveal metabolic complementarity between whiteflies and their symbionts. *Insect Sci.* 29, 539–549. doi: 10.1111/1744-7917.12943

frontiers | Frontiers in Plant Science

TYPE Review
PUBLISHED 21 February 2024
DOI 10.3389/fpls.2024.1279738

OPEN ACCESS

EDITED BY
Yinglong Chen,
University of Western Australia, Australia

REVIEWED BY
Vijay Sheri,
East Carolina University, United States
Thakku R. Ramkumar,
Delaware State University, United States

*CORRESPONDENCE
Jae-Yean Kim
kimjy@gnu.ac.kr
Moonhyuk Kwon
mkwon@gnu.ac.kr

RECEIVED 18 August 2023
ACCEPTED 02 February 2024
PUBLISHED 21 February 2024

CITATION
Das S, Kwon M and Kim J-Y (2024) Enhancement of specialized metabolites using CRISPR/Cas gene editing technology in medicinal plants.
Front. Plant Sci. 15:1279738.
doi: 10.3389/fpls.2024.1279738

Enhancement of specialized metabolites using CRISPR/Cas gene editing technology in medicinal plants

Swati Das[1], Moonhyuk Kwon [2*] and Jae-Yean Kim [1,3*]

[1]Division of Applied Life Science (BK21 Four Program), Plant Molecular Biology and Biotechnology Research Center (PMBBRC), Gyeongsang National University, Jinju, Republic of Korea, [2]Division of Life Science, Anti-aging Bio Cell Factory Regional Leading Research Center (ABC-RLRC), Research Institute of Molecular Alchemy (RIMA), Gyeongsang National University, Jinju, Republic of Korea, [3]Nulla Bio R&D Center, Nulla Bio Inc., Jinju, Republic of Korea

Plants are the richest source of specialized metabolites. The specialized metabolites offer a variety of physiological benefits and many adaptive evolutionary advantages and frequently linked to plant defense mechanisms. Medicinal plants are a vital source of nutrition and active pharmaceutical agents. The production of valuable specialized metabolites and bioactive compounds has increased with the improvement of transgenic techniques like gene silencing and gene overexpression. These techniques are beneficial for decreasing production costs and increasing nutritional value. Utilizing biotechnological applications to enhance specialized metabolites in medicinal plants needs characterization and identification of genes within an elucidated pathway. The breakthrough and advancement of CRISPR/Cas-based gene editing in improving the production of specific metabolites in medicinal plants have gained significant importance in contemporary times. This article imparts a comprehensive recapitulation of the latest advancements made in the implementation of CRISPR-gene editing techniques for the purpose of augmenting specific metabolites in medicinal plants. We also provide further insights and perspectives for improving metabolic engineering scenarios in medicinal plants.

KEYWORDS

CRISPR/Cas, gene editing, metabolic engineering, specialized metabolites, medicinal plants

1 Introduction

Plants with therapeutic and pharmacological importance produce great natural products applicable for human healthcare with a high market value in the areas of medicine, antioxidants, essences, perfumes, dyes, insecticides, pheromones, and other high-value natural goods for human healthcare (Yuan et al., 2016). Plants have developed

their metabolic systems in response to environmental challenges, resulting a diverse array of specialized metabolites production (Springob and Kutchan, 2009).

Plant-synthesized endogenous organic compounds are classified into two main classifications, namely primary and secondary (specialized) metabolites (Sangwan et al., 2018). In order for organisms to ensure their survival, it is imperative that they must possess essential metabolites, including carbohydrates, proteins, lipids, and nucleic acids (Sangwan and Sangwan, 2014b). Precursors for the specialized metabolites are the primary metabolites. The specialized metabolites confer many adaptive and evolutionary advantages with various metabolic functions and are often associated with plant defense mechanisms (Trethewey, 2004). Plants are the richest source of specialized metabolites, with an estimated 100,000 specialized metabolites in about 50,000 plant species (Pyne et al., 2019; Kaushik et al., 2017; Hussein and El-Anssary, 2018; Yang et al., 2018). Plant-specialized metabolites can be categorized into three primary classifications: terpenes, nitrogen-containing compounds, and phenolics, based on their biosynthetic pathways, and are categorized into volatiles, sterols, carotenoids, glycosides, alkaloids, glucosinolates, and tannins (Bourgaud et al., 2001; Jadaun et al., 2017). Specialized metabolites play a pivotal role in facilitating plants to combat adverse environmental and physiological stresses, unlike the primary metabolites which are essential for plant growth and development (Verpoorte et al., 2002; Singer et al., 2003).

Plant-specialized metabolites are essential for health improvement; they are used for antitumor, antimicrobial, anticancer, and antibiotic, such as bleomycin, pentacyclic terpenearjunolic acid, taxol, and penicillin, respectively (Sangwan et al., 2018). Medicinally significant specialized metabolites found in plants like *Taxus*, *Withania*, *Centella*, *Artemisia*, and *Cymbopogon* include taxol, artemisinin, withaferin, asiaticoside, withanone, madecassoside and essential oil (Sangwan et al., 2004, 2007; Sangwan et al., 2008; Chaurasiya et al., 2012; Sangwan et al., 2014a; Sangwan and Sangwan, 2014b). Traditional uses for these therapeutic herbs date back thousands of years.

In the contemporary age of biotechnology, research on medicinal plant-based bioactive compounds constantly upgrades the information in databases. Also, contributes to the growth of therapeutic drugs, agro-food, agrochemical, and beauty (Hassan, 2012). *Salvia miltiorrhiza*, *Cannabis sativa, Dendrobium officinale*, and *Opium poppy* have effective transformation mechanisms and high-quality reference genomes (Xu et al., 2016; Guo et al., 2018; Gao et al., 2020; Niu et al., 2021). Scientists are finding novel synthetic methods and focusing on mining critical genes for functional gene studies, producing effective compounds, and metabolic network regulation of medicinal plants for improving the quality and breeding superior germplasm. Retrieving genomic data from an ever-expanding variety of plant species and manipulating the genome is now possible thanks to the advancement of better genome engineering systems. These systems allow for precise and effective editing of target-specific genes at predetermined loci within a genome, repurposing the functions of specific targeted regions Figure 1A.

Targeted genome editing works by using a sequence-specific nuclease to instigate a DNA double-strand break (DSB) at a target site that repairs by either the donor-dependent homology-directed repair (HDR) pathway or the erroneous non-homologous end joining (NHEJ) pathway repair (Zhu et al., 2020). Early generation sequence-specific nucleases that depend on protein-DNA interactions, such as meganuclease (MN), zinc-finger nucleases (ZFNs), and transcriptional activator-like effector nucleases (TALENs), have adequately contributed to plant gene editing (Puchta et al., 1993; Wright et al., 2004; Christian et al., 2010). However, the impediment arises with their construction because their production necessitates intricate protein engineering. For a pre-determined sequence, nucleic acid cleavage activity performed by the third generation of CRISPR (clustered regularly interspaced short palindromic repeats)/Cas (CRISPR-associated protein) relying on DNA-RNA recognition demonstrates a greater advantage than protein-DNA interaction requiring complex engineered protein (Mali et al., 2013). Thus, CRISPR/Cas9 and its other orthologs or alternative CRISPR/Cas systems can introduce DSBs at specified target sites with little effort and expense. The DNA modifications the genome editing tools induce could be deletions, insertions, or substitutions (Bhatta and Malla, 2020). Due to its affordability compared to herbal medicines and allopathy and synthetic drugs, the global market for plant-derived products is anticipated to gain momentum over the upcoming years with the advancement of modern biotechnology. The growth in the plant-derived medicine market is due to increasing demand and intensive research investments and funding, and hopefully, it will increase rapidly within the next few decades.

This review paper examines the utilization of CRISPR/Cas approach to augment the biosynthesis of specific metabolites in agriculturally significant crops, including wheat, tomato, and rice. We also emphasize medicinal plants that are enriched by natural bioactive compounds necessary for industrial and pharmaceutical purposes, for instance, medical drugs, perfume, and cosmetic industries (Hassan, 2012). Subsequently, our focus shifts towards the plausible application of gene editing techniques for enhancing the biosynthesis of certain natural compounds *in vitro*.

2 Utilization of CRISPR/Cas gene editing approach in the plant biosynthetic pathway

2.1 The CRISPR/Cas system and how it operates

An adaptive system for phage immunity in bacteria and archaea, CRISPR/Cas, originates from prokaryotes. CRISPR/Cas system allows programmable RNA-guided genetic manipulation, as it is dependent on the recognition of DNA-RNA interaction and pre-determined sequence-specific nucleic acid cleavage binding activity (Zhu et al., 2020). This third-generation genome editing technology has the advantage over the earlier gene editing systems, namely, ZFN (Wright et al., 2004) and TALEN (Christian et al., 2010), because

their fabrication requires complex protein engineering, high cost, and lack of versatility for building a multiplex mutation system (Knott and Doudna, 2018; Kim and Kim, 2019; Chen et al., 2019a). The single-guide RNA (sgRNA), which comprises the fusion of a trans-activating RNA (tracrRNA) and a CRISPR RNA (crRNA), is part of the CRISPR/Cas system, derived from type II *Sp*Cas9 (the most widely adopted Cas9 protein from *Streptococcus pyogenes*) and has extensive utilization in various species, including plants. The sgRNA-Cas combined complex binds to the desired DNA sequence, cleaves the DNA, and recognizes a protospacer adjacent motif (PAM) site adjacent to the sgRNA binding site (Jinek et al., 2012). Another popular Cas enzyme, Cas12a, is a class 2, type V, RNA-guided DNA endonuclease having a single nuclease domain that requires a single, smaller gRNA molecule, RNAase activity for processing crRNA, generates staggered DNA ends, and detects T-rich PAM sites (Kadam et al., 2018).

The fundamental mechanism underlying the CRISPR/Cas tool is to achieve DSB at a designated genomic locus and repair broken gaps in DSBs by introducing DNA modifications by the donor-dependent HDR pathway or the erroneous NHEJ pathway (Guo et al., 2022). By reconnecting the two ends of the DSB, erroneous NHEJ introduces inaccurate short nucleotide insertions and deletions (indels) into the targeted regions (such as genes and promoters), causing frameshift alterations that interfere with the original structure and function (Jacobs et al., 2015). In juxtaposition to DSBs repaired by NHEJ, the repairs facilitated by HDR involve the precise insertion or replacement of a predefined nucleotide sequence using external homologous donor templates (Figure 1F).

Due to the expeditious progression of the CRISPR/Cas technology, it has undergone significant developments that surpass traditional targeted mutagenesis mediated by DSBs. These advancements encompass various gene editing tools such as base editors (Kim, 2018), high precision editing achieved through HDR and prime editors (Anzalone et al., 2019; Huang and Puchta, 2019), as well as transcriptional regulation (Mahas et al., 2018). Furthermore, the CRISPR/Cas system possesses the capacity to alter the expression of genes through the manipulation of transcriptional regulatory elements, including promoters, enhancers, terminators, and other related components (Zhang et al., 2019b).

Base editing and prime editing approaches do not rely on conventional CRISPR/Cas mechanisms for repairing a DSB in the gene sequence. The base editing tools are classified into cytosine base editors (CBEs) and adenine base editors (ABEs). These base editing tools comprise either a nicked Cas9 (nCas9) or catalytically inactivated Cas9 (dCas9) coupled with a particular deaminase. The deaminases induce transitory alterations in DNA by converting C•G to T•A or A •T to G •C, depending on their unique functions (Komor et al., 2016). Moreover, the prime editing tools comprised two main components: a fusion protein that includes Cas9 nickase (H840A) and reverse transcriptase, in addition to a second component called prime editing guide RNA (pegRNA). The Cas9 nickase variant (H840), consisting of a RuvC functioning domain, introduces nick to the non-target DNA strand. Reverse transcriptase then works with a pegRNA template to modify the necessary DNA (Anzalone et al., 2019). The primer editor tool has

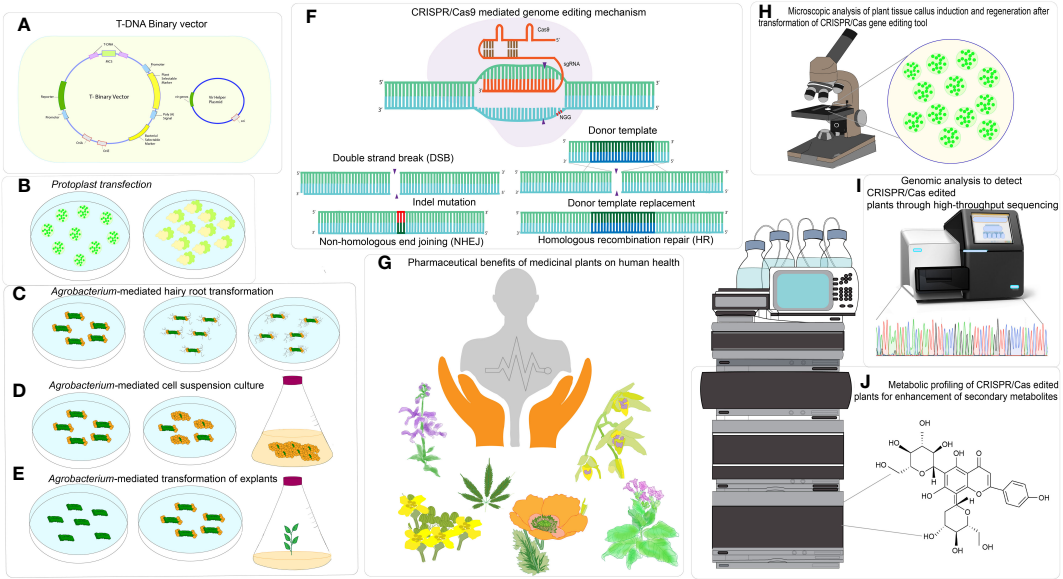


FIGURE 1
CRISPR/Cas based genome editing in medicinal plants for enhancement of specialized metabolites. **(A)** T-DNA Binary vector, **(B)** Particle bombardment of plasmid or CRISPR ribonucleoprotein (RNP) in protoplast, **(C)** *Agrobacterium*-mediated hairy root transformation, **(D)** *Agrobacterium*-mediated cell suspension culture, **(E)** *Agrobacterium*- mediated transformation of explants, **(F)** CRISPR/Cas-mediated genome editing mechanism, **(G)** Pharmaceutical benefits of medicinal plants on human health, **(H)** Microscopic analysis for analysis plant tissue callus induction and regeneration after transformation of CRISPR/Cas tool delivery, **(I)** Genomic analysis to detect CRISPR-edited plants through High throughput sequencing, **(J)** Metabolic profiling of CRISPR- edited plants for enhancement of secondary metabolites.

evolved into a sophisticated platform for facilitating accurate base substitution, DNA deletion, and DNA insertion.

The CRISPR/Cas system also involves in gene regulation via two main methods: transcriptional activation, referred to as CRISPRa, and transcriptional repression, known as CRISPR interference or CRISPRi (Qi et al., 2013; Chavez et al., 2015; Konermann et al., 2015). Moreover, recent advancements have resulted in innovative CRISPR-based methodologies aiming to fuse of epigenetic modifiers with dCas9. These strategies have been undertaken to modify the epigenome to attain precise alterations in DNA methylation and histone modifications (Abudayyeh et al., 2017; Gootenberg et al., 2017; Zhang et al., 2019b).

2.2 Current status of plant metabolic engineering in CRISPR era

The methodologies employed in the domain of metabolic engineering in plants can be categorized into two primary objectives: (1) nutritional quality improvement or biofortification, and (2) plant synthetic biology (Camerlengo et al., 2022). Enhancing the nutritional composition of food, often known as biofortification, is a critical objective in applied plant biology research. According to the World Health Organization (UN Report, 2022), 828 million individuals across the globe, particularly in underdeveloped and emerging nations, experience malnutrition due to inadequate nutrient intake. Considering the profound influence of nutrition on human welfare, implementing the CRISPR/Cas tool in biotechnology exhibits the potential for augmenting the nutritional profile of crop plants (Mora-Vásquez et al., 2022). Recently, several desired outcomes in crop plants engineered for enhancing specialized metabolites have been standardized and accustomed through targeted alteration of essential biosynthetic or metabolic processes within the genome, as outlined in Table 1. The understanding of current knowledge in CRISPR/Cas-mediated plant metabolic engineering in crop plants serves as a prototype for strategies applicable to increasing medicinally-valued metabolites in medicinal plant with limited information.

Rice and tomato have been enriched with carotenoid and lycopene accumulation by targeting the carotenoid biosynthetic pathway using CRISPR/Cas-mediated multiplexing knockout by NHEJ repair, and targeted homologous recombination inducing DSBs, or targeted insertion of DNA fragments without selectable markers (Dong et al., 2020; Li, X. et al., 2018; Globus and Qimron, 2018; Hayut et al., 2017; Shlush et al., 2021).

CRISPR/Cas9 multiplex genome editing utilized to alter five genes from the lycopene metabolic pathway, namely lycopene E-cyclase (*LCY-E*), *SGR1*, lycopene β-cyclase 1, β-lycopene cyclase (*Blc*), and *LCY-B2*. It reported that the lycopene concentration in tomato fruit increased by a factor of 5.1 after implementing multiplex genome editing (Li et al., 2018). Moreover, CRISPR/Cas9, was also employed to induce targeted homologous recombination (HR) in somatic cells, aimed at investigating the involvement of two particular genes, carotenoid isomerase (CRTISO) and phytoene synthase (PSY1), in the carotenoid biosynthesis pathway of tomato plants (Filler Hayut, Melamed Bessudo, and Levy 2017; Shlush et al., 2021). Another study in rice plants exhibited enhanced carotenoid levels in the seeds effectively produced by a targeted insertion of a 5.2 kb-sized carotenoid biosynthetic cassette without a selectable marker into the rice genome. The 5.2 kB-sized cassette had the coding sequences of two genes, specifically *SSU-crt*I and *ZmPsy* (Paine et al., 2005; Dong et al., 2020).

Another noteworthy example is the CRISPR/Cas9-mediated genetic alteration to increase the levels of the specialized metabolite, GABA (γ-aminobutyric acid), a non-proteinogenic amino acid, in prominent crops, such as rice and tomato. GABA, known for its hypotensive properties, with increased GABA levels in tomato fruits or rice grains, has the potential to confer advantageous effects on human health by reducing blood pressure. In their study, Nonaka et al. (2017) and Akama et al. (2020) focused on introducing a premature stop codon in close proximity to the autoinhibitory Ca2+/calmodulin binding domain (CaMBD) of glutamate decarboxylase (GAD) genes to enhance the GABA accumulation in tomato fruit, and rice, respectively (Nonaka et al., 2017; Akama et al., 2020). Li et al. (2018) devised an alternative way to strategize a multiplex CRISPR/Cas strategy enabling efficient targeting of five distinct genes associated with the GABA shunt pathway, which included GABA-TP1, GABA-TP2, GABA-TP3 (also referred to as TP-pyruvate-dependent GABA transaminase), CAT9, and SSADH (Koike et al., 2013; Snowden et al., 2013; Bao et al., 2015; Li et al., 2018).

The prevalence of nutritional deficiencies, such as vitamin D deficiency, is a significant concern on a global scale. Deficiency of vitamin D is associated with increased susceptibility to cancer, neurocognitive decline, and impaired skeletal development. Li et al. (2022) successfully attempted to increase 7-dehydrocholesterol (7-DHC) levels, often referred to as provitamin D3, in tomatoes, effectively with the employment of CRISPR/Cas9. A supplementary metabolic pathway related to steroidal glycoalkaloids (SGA) biosynthesis in tomato, wherein a distinct isoform known as 7-dehydrocholesterol reductase (Sl7-DR2) facilitates the conversion of 7-DHC to cholesterol, which is essential for the production of α-tomatine in both tomato leaves as well as fruit. To enhance the provitamin D3 levels in tomatoes through biofortification, the activity of Sl7-DR2 was disrupted, leading to the buildup of 7-DHC (Jäpelt and Jakobsen, 2013; Sonawane et al., 2016; Li, J. et al., 2022).

The development of purple pigmentation in plant tissues is greatly influenced by anthocyanin accumulation, a unique plant metabolite closely linked to enhanced resistance to herbivory, fungal pathogens, bacterial infections, and heavy metal-induced stressors. Anthocyanin, a group of water-soluble flavonoids, plays a significant role in distinct physiological processes, namely gastrointestinal absorption, bowel function, and cardiovascular and neurological diseases (Khusnutdinov et al., 2021). R2R3MYB predominantly controls anthocyanin biosynthesis regulation, the

TABLE 1 Application of CRISPR/Cas gene editing technology for enhancement of specialize metabolites in plants.

Plant species	Target gene	Nutritional/ Pharmaceutical Improvement	Targeted metabolic Pathway	Strategy	Result	Reference
Tomato	LCY-E, SGR1, Blc, LCY-B1, LCY-B2	Vitamin A enhancement; Lycopene accumulation	CRISPR/Cas9 mediated Knockout Carotenoid metabolic pathway	A bidirectional approach was adapted with the aim of augmenting lycopene production while concurrently inhibiting the conversion of β- and α-carotene from lycopene.	Lycopene concentration in tomato fruit increased by a fold of 5.1, subsequent after implementing multiplex genome editing	(Li, X. et al., 2018)
Rice	CrtI and PSY	Vitamin A enhancement; beta- carotene	CRISPR/Cas9 mediated 5.2kB sized insertion without selectable marker. Carotenoid metabolic pathway	The 5.2 kB-sized cassette had the coding sequences of two genes, specifically *SSU-crtI* and *ZmPsy*	Increased carotenoid levels in the rice seeds	(Paine et al., 2005; Dong et al., 2020, 2020)
Tomato	SlGAD2, SlGAD3, SlGABA-Ts, SlCAT9 and SlSSADH	GABA accumulation	CRISPR/Cas9-mediated C-terminal deletion of Calmodulin domain of GABA shunt pathway	Introducing pre-mature stop codon and gene disruption.	Increased GABA content in tomato fruit and leaves.	(Nonaka et al., 2017; Li, R. et al., 2018)
Rice	OsGAD3	GABA accumulation	CRISPR Cas9 mediated C-terminal deletion of Calmodulin domain of GABA shunt pathway	Introducing pre-mature stop codon	Increased GABA content in rice seeds	(Akama et al., 2020)
Tomato	7-dehydrocholesterol (7- DHC)	Enriched Vitamin D by duplicate SGA biosynthetic pathway	CRISPR/Cas9-mediated knockout of SGA biosynthesis	7-dehydrocholesterol reductase (Sl7-DR2) facilitates the conversion of 7-DHC to cholesterol, which is essential for the production of α-tomatine in both tomato leaves as well as fruit	. To enhance the provitamin D3 levels in tomatoes through biofortification, the activity of Sl-DR2 was disrupted, leading to the buildup of 7-DHC	(Li, J. et al., 2022)
Tomato	R2R3MYB, Anthocyanin Pigment 1 (PAP1, MYB75), DFR, F3H, and F3'H and genes *ANT1*, *AN2*-like, *AN2*	Anthocyanin accumulation	CRISPR/Cas9-mediated TF gene knockout	TFs family member gene targeting	Study of functional role of TFs in anthocyanin biosynthesis pathway	Riaz et al., 2019; Wang et al., 2019
Lettuce	LsGGP2 uORF	Improvement of antioxidants, oxidative stress tolerance and 150% increase in ascorbate content	CRISPR/Cas9 mediated Knockout of SGA biosynthesis			(Si et al., 2020; Zhang, et al., 2018b)
Lettuce	CPT3	Natural Rubber	Rubber biosynthesis			(Kwon et al., 2023)
Brassica napus	BnITPK	Free of Phytic acid	35% decrease in phytic acid by CRISPR Cas9 mediated gene editing			(Sparvoli and Cominelli, 2015)
Wheat	TaIPK1	Free of Phytic acid	Reduced levels of Phytic acid, increase Fe 1.5-2.1 fold to 1.6 to 1.9 fold			(Sashidhar et al., 2020)
Wheat	TaASN2	Free of Asparagine	90% Reduction in asparagine			(Raffan et al., 2021)

Anthocyanin Pigment 1 (PAP1, MYB75) gene family, DFR, F3H, and F3'H and genes *ANT1*, *AN2*-like, *AN2* (unique to *Solanum lycopersicum*), making them desirable target for CRISPR/Cas gene editing for anthocyanin accumulation (Riaz et al., 2019; Wang et al., 2019).

The manipulation of endogenous upstream open reading frames (uORF) inside plant genomes allows for controlling the regulation of mRNA translation originating from primary open reading frames (ORFs). According to Zhang et al. (2018a) and Si et al. (2020), the editing of uORFs, in the *LsGGP2* gene, a crucial enzyme in the production of Vitamin C, led to a significant enhancement in oxidative stress tolerance and ascorbate content, with a notable rise of 150% (Zhang et al., 2018b; Si et al., 2020).

The prospect of utilizing CRISPR/Cas9-mediated gene editing in plant metabolic engineering expands and unfolds mechanisms of novel biosynthetic pathways in plants. For instance, the knockout variants of *LsCPT3*, Kwon et al. (2023) optimized by the CRISPR/Cas9 tool in lettuce plants. In lettuce laticifers that have undergone CPT transformation, it is possible to make natural rubber with a molecular weight (Mw) exceeding 1 million Daltons. Conversely, it noted that indigenous golden rods could produce natural rubber (NR) with a molecular weight of approximately 0.1 million Daltons (Da). This characteristic provides significant insights into the mechanism of biosynthesis of natural rubber in plant life (Kwon et al., 2023). Recent advancements in the metabolic engineering of agricultural plants have facilitated the augmentation of specialized metabolite levels that have provided the discovery of novel roles within biosynthetic pathways and the enhancement of nutritional or commercial value in these plants.

3 Utilizing CRISPR/Cas gene technology in medicinal plants to increase the production of specialty metabolites

Medicinal plants provide bioactive specialized metabolites serving as valuable derivatives for the commercial therapeutics industry Figure 1G. Improvements in the synthesis of these medicinally valuable specialized metabolites obtained in the decades before the CRISPR/Cas era. The non-CRISPR/Cas technology involved in genetic manipulation by application of transformation techniques, including gene silencing, gene stacking, and overexpression (Wilson and Roberts, 2014), with a concurrent aim to reduce production expenses, although not as precise manipulation as the CRISPR/Cas system. The application of metabolic engineering for augmenting the synthesis of specialized metabolites and other natural products in medicinal plants, utilizing non-CRISPR/Cas genetic manipulation and CRISPR/Cas-mediated genome editing technology, is exemplified in Tables 2, 3.

3.1 Current status of enhancement of specialized metabolites through non-CRISPR/Cas technology in medicinal plants

Utilizing biotechnological applications to enhance specialized metabolites in medicinal plants needs the characterization and identification of genes regulating an elucidated pathway. It is

TABLE 2 Application of non-CRISPR/Cas technology for the enhancement of specialized metabolites in medicinal plants.

Plant species	Gene manipulation technique	Specialized metabolites	Target gene/Purpose	Reference
Papaver bracteatum	Endogenous expression of CodR, a crucial enzyme for converting thebaine to morphine	Elevated codeine levels by eleven-fold and morphine levels with marginal increase of 0.28 percent	Upregulated codeinone reductase (*CodR*)	Sharafi et al., 2013
Scopolia lurida	Introduction of external foreign enzyme	Accumulation of scopolamine	Overexpression of hyoscyamine 6β-hydroxylase (*Hn*H6H) from *Hyocyamus niger*	Lan et al., 2018
Catharanthus roseus	Overexpression of transcription factor of a specific metabolic pathway	Elevated tabersonine, an indole alkaloid for vinblastine production	Overexpression of *ORCA4* TF	Paul et al., 2017
Artemisia annua	RNAi	*AaC4H* RNAi lines exhibited an elevation of trans-cinnamic acid content and reduction in p-coumaric acid levels, salicylic acid (SA) and artemisinin	Gene silencing of cinnamate-4-hydroxylase (*AaC4H*) responsible for the conversion of trans-cinnamic acid (CA) to p-coumaric acid (COA)	Kumar et al., 2016
Artemisia annua	RNAi	Accumulation of alkaloid, salutaridine	Suppression of specific gene responsible for expressing the enzyme salutaridinol 7-O-acetyltransferase (*SalAT*)	Allen et al., 2008
Artemisia annua	RNAi	Enhancing artemisinin level by 3.14-fold	Gene silencing of squalene synthase (*SQS*), that competes with artemisinin pathway	Zhang et al., 2009
Catharanthus roseus	RNA-sequencing	Construction of a database 'CathaCyc', with metabolic pathway-related information	Facilitate visualization and interpretation of transcriptomic data	Strickler et al., 2012
Ophiorrhiza pumila	Metabolomics	Metabolomic analysis of suspension culture and hairy roots	Distinguish between potential genes regulating monoterpene indole alkaloids and anthraquinones	Higashi and Saito, 2013

TABLE 3 Application of CRISPR/Cas gene editing technology for enhancement of specialized metabolites.

Plant species	Specialized metabolites	Target gene	CRISPR/Cas gene editing strategy	Target site; sgRNA and Cas9 Promoter	Transformation method	Mutation Rate; Gene-editing improvement	Reference
Salvia miltiorhiza	Laccase (multi-copper containing glycoproteins)	SmLACs (SmLAC7 and SmLAC20)	Targeting conserved domains to knockout multiple genes of laccase family; functional study	First exon; AtU6::sgRNA; CaMV35S:: SpCas9	Hairy roots	Accumulation of lignin and phenolic acid	(Zhou et al., 2021)
Salvia miltiorhiza	Phenolics compounds	bZIP	Knockout lines; functional study to prove negative regulator	First exon sequence od bZIP2; AtU6:: sgRNA; CaMV35S:: hSpCas9	Hairy roots	40% mutation rate, elevated phenolic acid, PAL (phenylalanine ammonia-lyase)	(Shi et al., 2021)
Salvia miltiorhiza	Phenolics compounds RA and LAB	RAS	Knockout	ORF;AtU6:: sgRNA; OsU3:: sgRNA CaMV35S:: hSpCas9	Hairy roots	50% in AtU6::sgNA, Lithospermic androsmaniric acid content were reduced	(Zhou et al., 2018)
Salvia miltiorhiza	Tanshinones	CPS1	Multiplex knockout	First,fourth, eleventh exon. AtU6::sgRNA1, sgRNA2, sgRNA3; CaMV35S:: hSpCas9	Hairy roots		(Li et al., 2018)
Atropa belladona	Anticholinergic tropane alkaloids (TAs), hyoscyamine, scopolamine anisodamine.	AbH6H	Pre-mature stop codon	Second exon of ORF;AtU6:: sgRNA; CaMV35S:: SpCas9	Tissue culture cotyledon	63.6% Mutation rate;3.68-4.21 Fold increase in Hyoscyamine	(Zeng et al., 2021)
Dendrobium officinale	Polysaccharides, bibenzyls, essential oil	C3H, C4H, 4CL, CCR and IRX	Frameshift	OsU3::sgRNA; CaMV35S:: pcoCas9	Hairy roots	10%-100% editing efficiency	(Kui et al., 2017)
Nicotiana tabacum	Plant-derived glycoprotein immune-responsive residue	XylT and FucT	Multiplex knockout	1st exon and 3rd exon; U6:: sgRNA, 35S::pcoCas9	Tissue culture	No detectable FucT and XylT residue.	(Mercx et al., 2017)
Nicotiana tabacum	Nicotine-free	BBl	Knockout	AtU6::sgRNA; 35S::SpCas9	Tissue culture	99.6% Nicotine free	(Schachtsiek and Stehle, 2019)
Cannabis sativa	THC-free (proposed)	CsPDS; THCA	Multiplex Knockout	Exon6, multiple sgRNA	Tissue culture	Albino phenotype	(Zhang et al., 2021)
Comfrey (*Symphytum officinale*)	Homospermidine toxic free	HSS	Knockout	AtU6::sgRNA; 35S::Cas9	Hairy roots	Reduced level of homospermidine	(Zakaria et al., 2021)
Opium Poppy (*Papaver somniferum*)	Benzylisoquinoline	4'OMT2	Knockout	AtU6::sgRNA; 35S::hCas9	Agroinfiltration	S- reticuline for benzylisoquinoline alkaloids production	(Alagoz et al., 2016)
Dioscorea zingiberensis	Diosgenin	Dzfps	Frameshift	First exon; OsU3::sgRNA; 35S::SpCas9	Hairy roots	Decreased squalene level	(Feng et al., 2018)
Taraxacum kok-saghyz	Natural rubber	1-FFT	Knockout	AtU6::sgRNA; 35S::pcoCas9	Hairy root	Natural rubber synthesis	(Iaffaldano et al., 2016)

crucial to consider the interplay of components within the metabolic pathways during specific specialized metabolite biosynthesis, including its interactions with enzymes, gene regulation, sub-cellular localization, and epigenetic regulation (Mora-Vásquez et al., 2022). Advanced biotechnological approaches, such as proteomics and functional genomics, help decipher genes that encode their corresponding regulatory enzymes. The term "gene manipulation strategies" refers to a wide range of methods, including the up-or down-regulation of particular metabolic pathway genes or the enzymes responsible for controlling the rate-limiting steps, multi-gene stacking within the same chromosomal vicinity, targeting or blocking of branching pathways that interfere with the biosynthesis of the desired active product, alteration of transcription factor expression levels that directly or indirectly regulate multiple genes within a biosynthetic pathway, and introducing foreign genes. Therefore, enhancing the accumulation of medicinally valuable specialized metabolites can be achieved by increasing the abundance of these metabolites within the plant system or by augmenting the metabolic flux of a particular pathway using advanced molecular tools (Sanchez and Demain, 2008).

The non-CRISPR/Cas genetic engineering techniques, which involve augmenting the gene expression of a catalyzing enzyme, precursor, or product within a metabolic pathway, can be bifurcated into three distinct methods. To begin with, the endogenous expression of crucial enzymes that regulate the rate of various intricate reactions in a pathway, such as methylation, condensation, isomerization, glycosylation, and acylation, has the potential for an augmented buildup of specialized metabolite accumulation in the plant system (Kulkarni, 2016). One example is *Papaver bracteatum*, a botanical species known for its therapeutic properties characterized by a significant presence of thebaine, but exhibiting relatively lower levels of codeine and morphine. The upregulation of codeinone reductase (*CodR*), the pivotal enzyme for converting thebaine to morphine, leads to a significant elevation in codeine levels by eleven-fold and a marginal increase of 0.28 percent D.W. in morphine content in genetically-engineered hairy roots (Sharafi et al., 2013). Secondly, an alternative strategy is introducing an external foreign enzyme derived from a distinct plant species, which may demonstrate enhanced catalytic efficacy in the conversion process toward the targeted specialized metabolite. The overexpression of the exogenous gene hyoscyamine 6β-hydroxylase (*Hn*H6H) from *Hyocyamus niger* in hairy root cultures of *Scopolia lurida* led to a substantial ten-fold enhancement in the accumulation of scopolamine in comparison to the endogenous overexpression of the gene *Sl*H6H (Lan et al., 2018). The third method is to overexpress any transcription factor regulating the gene transcription of a specific metabolic pathway. The upregulation of one transcription factor has several probabilities to modify multiple interconnected genes of a biosynthetic pathway; also cis-regulatory elements that interact with the DNA-binding domain of transcription factors can have an impact on activating or suppressing related target genes (Mahjoub et al., 2009; Samad et al., 2017). Out of five transcription factors belonging to AP2/ERF TFs in *Catharanthus roseus*, namely, *ORCA2* to *ORCA6* that regulates terpenoid indole alkaloids (TIAs), the upregulated genes associated with the indole and seco-iridoid pathways occur when the *ORCA4* TF is overexpressed and elevates the level of tabersonine, an indole alkaloid crucial for vinblastine production (Paul et al., 2017; Singh et al., 2020) (Table 2).

Another gene modification approach involves the accumulation of intermediate precursors in a metabolic pathway by silencing subsequent gene expression in that particular pathway (Sinha et al., 2019). Gene silence refers to inhibiting gene expression, which can occur at either the transcription or translation level. Various methods have been developed to achieve gene silencing, including RNA interference (RNAi), miRNA, and Virus-induced gene silencing (VIGS) (Singh et al., 2018; Li et al., 2015; Ossowski et al., 2008; Liu et al., 2017).

RNAi type of suppression or post-transcriptional gene silencing is an approach with which a particular mRNA gets degraded, associated with double-stranded DNA. Medicinal plants, for instance, *Artemisia annua*, *Catharanthus roseus*, *Nicotiana tabacum*, *Panax ginseng*, *Panax quinquefolius*, *Papaver somniferum*, *Withania somnifera* have been engineered with an RNAi approach (Sinha et al., 2019).to suppress the expression of a gene accountable for the degradation of a desired metabolite (Sinha et al., 2019). The gene silencing methods experimented on cinnamate-4-hydroxylase (*AaC4H*) in *Artemisia annua* are responsible for the conversion of trans-cinnamic acid (CA) to p-coumaric acid (COA) in the phenylpropanoid/lignin biosynthetic pathway exhibited an elevation of trans-cinnamic acid, salicylic acid (SA), artemisinin content and reduction in p-coumaric acid levels (Kumar et al., 2016). In their study published in 2008, Allen et al. documented the accumulation of a hitherto unidentified alkaloid, salutaridine, through the use of RNA interference (RNAi) to repress a specific gene responsible for expressing the enzyme salutaridinol 7-O-acetyltransferase (*SalAT*) (Allen et al., 2008). Suppression of critical genes from a competitive pathway can also accumulate a preferred specialized metabolite. Zhang et al. (2009) worked on enhancing the artemisinin level in *A. annua* by utilizing the hairpin-RNA-mediated RNAi method to modify squalene synthase (*SQS*), an adherent member of the sterol biosynthesis pathway that also competes with the artemisinin pathway (Zhang et al., 2009) (Table 2). RNAi works at the post-transcriptional level to silence the expression of a gene. However, it does not eliminate the function of the targeted gene, analogous to any gene editing tool that permits precision editing within the genome with minimal off-target effects (Barrangou et al., 2015; Alagoz et al., 2016; Kui et al., 2017). The first- and second-generation gene editing tools available in the pre-CRISPR period included MNs, ZFNs, and TALENs (Mani et al., 2005; Bogdanove and Voytas, 2011; Deng et al., 2012). However, these previously designed pre-CRISPR era genome editing tools have limitations in targeting specific genome sequences because they require two distinct protein hybrid configurations, which scarcely recognize the existing regions flanking the target DNA (Kim and Chandrasegaran, 1996; Li

et al., 2011). Thus, ZFN and TALENs are tedious to design, less specific, and time-consuming. The application of ZFN and TALEN for genome editing was more substantial in model crop plants than in medicinal plants.

High-throughput sequencing-based 'omics' approaches are crucial for identifying and characterizing of novel genes within an unrecognized pathway once a reference genome is accessible (Wilson and Roberts, 2014). This is in addition to exploring genome editing tools in metabolic engineering. Researchers can employ quantitative trait loci (QTL) and genome-wide association studies (GWAS) to ascertain the potential genes accountable for particular phenotypes. In instances where genomic data is lacking for a given species, transcriptomic analyses can be utilized to identify potential biosynthetic pathway genes using differential expression studies. For example, the authors utilized RNA-sequencing data from *Catharanthus roseus* to construct a database with metabolic pathway-related information called CathaCyc. This database was developed to facilitate the visualization and interpretation of intricate transcriptome data (Strickler et al., 2012). The incorporation of data derived from metabolomic and proteomic investigations further enhances comprehension of transcriptome studies. In a study by Higashi and Saito, 2013, metabolomic analysis was performed on both suspension cultures and hairy roots of *Ophiorrhiza pumila*. These cultures were shown to accumulate specialized metabolites (Higashi and Saito, 2013). Through this analysis, the researchers could distinguish between potential genes in regulating the biosynthesis pathway of monoterpene indole alkaloids and anthraquinones (Table 2). Next-generation sequencing (NGS) methods and inventive CRISPR/Cas-based breeding tools have emerged as superior alternatives to previously employed protein-engineered gene editing tools namely, ZFNs and TALENs. The CRISPR/Cas editing system shows prospect of being a highly practical toolkit for molecular engineering in medicinal plants, especially regarding the augmentation of the specialized metabolite profiles (Niazian, 2019) (Figures 1I, J).

3.2 Employment of CRISPR/Cas gene technology for enhancement of specialized metabolites in medicinal plants

In the contemporary period of CRISPR/Cas-mediated gene editing, there is a growing emphasis on utilizing and refining of this technology to augment the specific metabolite profiles in medicinal plants. The CRISPR/Cas9 editing tool has proven to be remarkably constructive in the genetic modification of both crops and prototype organisms. It can induce a wide variety of targeted genetic modifications, including frameshift and premature termination codon, gene substitution, site-specific single-gene knockout, fragment deletion, gene substitution, and targeted gene insertion. Characterizing the phenotypic effects of genetic mutants is a methodical way to establish the functional significance of a gene that regulates the synthesis of a specific metabolite in a biosynthetic pathway. Table 3 conceptualizes the application of CRISPR/Cas-mediated gene editing techniques in medicinal plants.

3.2.1 CRISPR/Cas-mediated targeted mutagenesis/knockout in medicinal plants

In medicinal plants, CRISPR/Cas9-based gene disruption techniques for targeted mutagenesis have been widely applied for the enrichment of pharmaceutically beneficial specialized metabolites or reducing the levels of toxic metabolites. For instance, the modification of the *hyoscyamine 6β- hydroxylase* (*AbH6H*) gene in *Atropa belladonna* L. generates edited plants that lack anisodamine and scopolamine compounds (Zeng et al., 2021). *Atropa belladonna* is a noteworthy botanical specimen that assumes a pivotal function synthesizing of anticholinergic tropane alkaloids (TAs), including hyoscyamine, scopolamine, and anisodamine. These compounds possess considerable commercial worth within the pharmaceutical sector (Gieger and Hesse, 1833, Lossen et al., 1864). Hyoscyamine is found to have therapeutic value in treating arrhythmias and organophosphate poisoning. On the other hand, anisodamine has shown promise in managing of gastrointestinal colic and vascular spasms. Additionally, scopolamine is found to be effective in alleviating symptoms of motion sickness. Hyoscyamine is a prominent constituent of TAs, while anisodamine and scopolamine, which are derived from hyoscyamine, are considered minor constituents (Viraliyur Yun and Yamada, 1992; Poupko et al., 2007; Ramaswami and Thangavelu, 2015). Therefore, due to their similar structures, hyoscyamine's independent separation and purification from its two derivatives have become challenging and cost-inefficient from *A. belladonna* raw extracts. In their 2021 publication, Zeng et al. provided evidence that a bifunctional dioxygenase, *hyoscyamine 6β-hydroxylase* (*H6H*), is responsible for the enzymatic conversion of hyoscyamine to anisodamine via 6β-hydroxylation, followed by the subsequent epoxidation of anisodamine to scopolamine (Hashimoto and Yamada, 1986; Zeng et al., 2021). The authors employed CRISPR/Cas9-targeted mutagenesis techniques to modify *H6H*, aiming to intensify hyoscyamine production levels while preventing the formation of its derivatives, anisodamine and scopolamine. The sgRNA designed to target at the second exon within the ORF of the *Ab*H6H gene, regulated by the *At*U6-26 promoter, and the *Sp*Cas9 protein from *Streptococcus pyogenes* by the Cauliflower mosaic virus (CaMV35S) promoter. Seven plant lines out of the eleven transgenic plants examined in this study showed editing events characterized by frameshifts in the ORF region and changes to most amino acids, indicating a cumulative mutation rate of 63.6 percent (seven out of eleven) for the Cas-*H6H* system. Moreover, within the subset of these seven edited lines, three of the transgenic plant lines were observed to be homozygous, accounting for 27.3 percent (three out of eleven) of the total. The use of High-Performance Liquid Chromatography (HPLC) for analyzing alkaloids demonstrated the identification of a peak associated with hyoscyamine. However, no peaks related to anisodamine and scopolamine were observed in plants where the AbH6H-function had been disturbed (Zeng et al., 2021).

Three lines of *AbH6H* homozygous mutant roots exhibited a significant increase in hyoscyamine content with fold changes of 3.68-, 4.21-, and 4.28, respectively, compared to wild-type roots. With the specific goal of altering five target genes, CRISPR/Cas-based targeted mutagenesis was applied to the species *Dendrobium*

officinale, targeting the enzymes *Coumarate 3-hydroxylase* (C3H), *Cinnamate 4-hydroxylase* (C4H), *4-coumarate coenzyme A ligase* (4CL), *Cinnamoyl coenzyme A reductase* (CCR) and *Irregular xylem5* (IRX) involved in the biosynthesis of lignocellulose. These enzymes are responsible for catalyzing various reactions within the pathway. It was observed that these enzymes can introduce alterations bordering on substitutions, insertions, or deletions in the generated products. The frequency of these edits can range from ten to a hundred percent. *Dendrobium officinale*, a highly regarded Chinese medicinal herb, possesses diverse pharmacological properties such as immunomodulation, anti-tumor effects, antioxidant activity, anti-fatigue properties, and renal protection, as demonstrated in experiments conducted on diabetic rats (Gao et al., 2002; Chen et al., 2014, 2021). Three sgRNAs were strategically designed at separate positions for every target gene. The sgRNAs in this study were governed by the OsU3 promoter, whereas the Cas9 enzyme, which incorporated plant-optimized codons, was controlled by the CaMV35S promoter. The mutagenesis targeting diverse gene candidates exhibited similar editing efficiency across all targets. One potential justification for this phenomenon could be attributed to variations in the chromatin structure at the specific genomic loci. Consequently, the accessibility of the Cas9 protein and the sgRNA is hindered in chromosome regions exhibiting a higher degree of compaction (Kui et al., 2017). To explore the potential applicability of CRISPR/Cas-based targeted mutagenesis in *Dioscorea zingiberensis*, a medicinal monocotyledon with a notable abundance of diosgenin in its rhizome was identified with the farnesyl pyrophosphate synthase gene (*Dzfps*) as a suitable target for the study. Diosgenin functions as a vital enzyme precursor in the synthesis of diosgenin, rendering it a significant subject of study in the context of numerous steroid hormone medications. Its pharmacological attributes encompass noteworthy potential in anti-inflammatory, anti-allergic, and anti-oxidant activities (Feng et al., 2018). The procedure involves the formation of E-isomer farnesyl pyrophosphate (*FPP*), preceding the facilitation of the stepwise condensation of dimethylallyl diphosphate (*DMAPP*) and geranyl diphosphate (*GPP*). The targeting sequence design for the sgRNA was aligned with the initial exon of the *Dzfps* gene. While the *Os*U3 promoter was responsible for regulating the expression of sgRNA, the CaMV35S promoter was the one that determined how much *Sp*Cas9 was produced. The observed target sites displayed five unique categories of changes, specifically deletions of 1, 4, 5, 8, and 10 base pairs. It is worth noting that all mutant versions exhibited chimeric mutations. All observed alterations led to frameshift mutations and premature termination codons. The observed reduction in squalene levels suggests that using CRISPR/Cas9 resulted in compromised functionality of *Dzfps*. The works, as mentioned earlier, provide sophisticated illustrations of merging CRISPR/Cas9 tool utility for conducting gene editing in medicinally-valued plants (Feng et al., 2018).

Knockout of genes by CRISPR/Cas9 often occurs through indel mutation, fragment deletion, domain, or large deletion generated by the NHEJ repair pathway, resulting in loss of function. This reverse genetics strategy is helpful for functional studies of a gene or transcription factor (TF) controlling any gene regulatory pathway. Knockout lines deleting important domains of a transcription factor leads to understanding whether it is a negative or positive regulator of a biosynthetic pathway. In *Salvia miltiorrhiza* knockout lines of the basic leucine zipper2 (bZIP2), gene was generated, precisely targeting the bZIP domain. The findings of this work indicate that bZIP2 exerts a suppressive influence on phenolic acid production. Shi et al. (2021) conducted a study on a novel bZIP member called bZIP2 in *Salvia miltiorrhiza* which also exhibited a robust response to the induction of abscisic acid (ABA) (Shi et al., 2021). Previous studies have provided limited evidence of the effectiveness of ABA in modulating the synthesis pathway of phenolic acids in *Salvia miltiorrhiza* hairy roots (Cui et al., 2012). Thus, the functional significance of bZIP family genes in the elevation of phenolic acid levels in *Salvia miltiorrhiza* is currently obscure due to minimal research on the regulating ABA-mediated modulation of phenolic acid synthesis.

Seventy bZIP TFs were examined using whole genome data from *S. miltiorrhiza*. These TFs were categorized into eleven sub-groups. Notably, the group-A bZIP TFs, specifically those belonging to the ABI-ABF-AREB subfamily, exhibited significant similarity to group-A members found in other plant species (Zhang et al., 2018a). The TFs as mentioned above are recognized as essential contributors to the ABA-dependent signaling cascade. Furthermore, Bhatnagar et al. (2017) made the observation that the group-A bZIP TFs exhibited four conserved regions (C1-C4) in association with a bZIP domain. The CRISPR/Cas9 vector harbors an *At*U6-sgRNA cassette engineered to target and disrupt the *bZIP2* gene. The CaMV35S promoter regulates this vector. A total of 45 transgenic hairy roots were produced, among which 18 samples exhibited insertions and deletions (indels) in the designated target region. The observed editing rate in these samples was determined to be 40 percent. The bZIP2 knockout lines displayed a significant elevation in phenolic acid content, with values ranging from 47.70 mg/g D.W. to 59.35 mg/g D.W. The observed increase exhibited a substantial escalation ranging from 23 to 53 percent compared to the unedited wild-type control (Shi et al., 2021). Among the various downstream genes examined, it was observed that the expression level of phenylalanine ammonia-lyase (PAL) exhibited the most notable increments in the knockout mutants. At the same time, while it decreased in the overexpression lines. The results connote that in *S. miltiorrhiza*, bZIP2 has a role as a suppressor in the phenolic acids biosynthetic pathway, specifically rosmarinic acid (RA) and salvianolic acid B (Sal B) (Shi et al., 2021).

An additional illustration of a CRISPR-based knockout strategy is implemented to impair the biosynthesis of nicotine selectively assisted by flavoproteins affiliated with the berberine bridge enzyme-like (BBL) family. Previous research undertaken by Kajikawa et al. (2011; 2017) has revealed that six genes (BBLa, BBLc, BBLd.2) have their origins in *Nicotiana sylvestris*, while BBLb, BBLd.1, and BBLe have their origins in *Nicotiana tomentosiformis*. The research aimed was to identify homologous genes in *Nicotiana tabacum* and design sgRNA targeting a specific sequence that shares similarity with these genes, except for the PAM site (Figure 2E) (Kajikawa et al., 2011, 2017). The sgRNA was generated under the regulation of the *At*U6-26 promoter using the transformation vector pCas9-TPC. The vector utilized in this study contained a

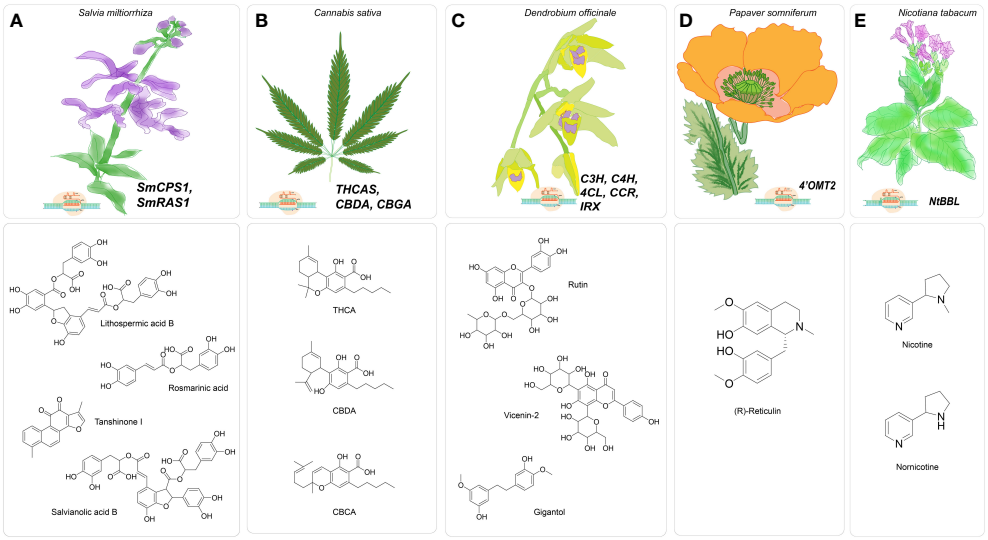


FIGURE 2
Major gene targets for CRISPR/Cas-mediated genome editing in medicinal plants **(A)**. CRISPR/Cas9-mediated genome editing in *Salvia miltiorrhiza* targeting *Sm*CPS1 and *Sm*RAS for enhancing specialized metabolites tanshinones, water-soluble phenolic acids, including rosmarinic acid, salvianolic acid B and lithospermic acid B; **(B)**. Proposed CRISPR/Cas9-based genome editing in *Cannabis sativa* targeting THCA, CBDA, CBGA to enhance cannabinoid production; **(C)**. CRISPR/Cas9- mediated gene editing in *Dendrobium officinale* targeting C3H, C4H, 4CL, CCR, IRX to enhance specialized metabolites bibenzyls of lignocellulosic pathway; **(D)**. CRISPR/Cas9-based genome *Papaver somniferum* targeting *Ps*4'OMT to improve synthesize of benzylisoquinoline, **(E).** CRISPR/Cas9-based genome editing in *Nicotiana tabacum* targeting *Nt*BBL for nicotine-free tobacco.

bar gene, which functioned as a marker for selection (Fauser et al., 2014). In the subsequent generations (T2 and T3), the nicotine level was determined to be 0.06 mg g per DW and 0.04 mg g per DW, respectively. The data mentioned above indicate a substantial decrease of 99.6 and 99.7 percent concerning to the wild-type. Notably, the T3 generation was determined to be largely lacking in nicotine (Schachtsiek and Stehle, 2019).

The CRISPR/Cas-based knockout technique was employed to investigate the regulation of benzylisoquinoline alkaloid (BIA) production in *Papaver somniferum*, a plant species known for its ability to biosynthesize morphine and BIAs utilized in the field of biomedicine (Figure 2D). The BIA biosynthesis commences with the amalgamation of dopamine and 4-hydroxyphenyl acetaldehyde (*4-HPAA*), resulting in the production of (S)-norcoclaurine (Fairbairn and Wassel, 1964; Valva et al., 1985; Ilari et al., 2009). The conversion of (S)-norcoclaurine into a key intermediate S-reticuline involves a sequence of methylation and hydroxylation reactions. The catalytic process is facilitated by 3'-hydroxy-N-methylcoclaurine 4'-O-methyltransferase (*4'OMT*) (Alagoz et al., 2016). Morphine, noscapine, and papaverine are final products that are obtained by distinct biosynthetic processes from S-reticuline. Previously, *4'OMT2* gene was overexpressed and subjected to TRV-mediated gene silencing experiments, but there is currently no conclusive evidence regarding its role in the synthesis of BIAs. Despite these studies, the biosynthetic pathway remains active, producing ongoing metabolite production. Consequently, the comprehensive phenotypic effects regulated by the targeted genes may be obscured (Gurkok et al., 2016). Therefore, knockout strategies utilizing a programmable CRISPR/Cas9 system effectively tackle the obstacles and elucidate the functions about the gene of interest. In this case, the 5'region of *4'OMT2* has been chosen as the target location for constructing a 20 base pair sgRNA. The viral expression vector system mediated by the Tobacco rattling virus (pTRV) was combined with the sgRNA, controlled by the AtU6 promoter, and the human-codon optimized version of Cas9 nuclease (hCas9), presided over by the CaMV35S promoter and nopaline synthase (nos) terminator. The resultant structure was injected into plant leaves via the agroinfiltration methodology. To assess the effectiveness of editing, the researchers utilized knockout lines to evaluate various compounds (morphine, S-reticuline, codeine, laudanosine, thebaine, noscapine, and papaverine) through the application of HPLC-ToF/MS and found their levels significantly reduced. The alkaloid accumulation, mostly thebaine and S-reticuline showed decreased levels for synthetic and viral-based CRISPR knockout lines (Alagoz et al., 2016).

In order to expedite the process of domesticating *Taraxacum kok-saghyz* (TK), commonly known as Rubber dandelion, a genetic element responsible for the production of *fructan-fructan 1-fructosyltransferase* (*1-FFT*) was targeted for disruption using the CRISPR/Cas9 approach. This gene contributes to the biosynthesis of inulin, a substance known to impair the production of rubber. *Taraxacum kok-saghyz* is recognized for its inherent capacity to synthesize rubber with a substantial molecular weight in its subterranean structures, rendering it a viable substitute for natural rubber. The selected site of the target for constructing sgRNA driven by the *At*U6-26 promoter was the second exon of the *1-FFT* gene due to its shared homology in all expected isoforms. Out of the eleven hairy roots, a significant proportion of ten were subjected to genome editing, demonstrating a mutation efficiency of 88.9%. According to Iaffaldano et al. (2016), it was observed that plants with genetically modified genomes were successfully regenerated from hairy roots in a time frame of six weeks (Iaffaldano et al., 2016).

One noteworthy report was conducted by Zhou et al. in 2018, wherein the researchers sought to clarify the functional importance of the rosmarinic acid synthase gene (*Sm*RAS) in the production of phenolic acids using the CRISPR/Cas9 approach (Zhou et al., 2018) (Figure 2A). A specific genomic locus, SMil_00025190, out of eleven homologous members in the RAS gene family, was selected from the *S. miltiorrhiza* genome database due to its notable expression across multiple organs. The sgRNA was strategized accordingly, to specifically binds with the targeted ORF of *Sm*RAS. The sgRNA utilized in this study was regulated by endogenous plant promoters, namely the *At*U6-26 derived from *Arabidopsis thaliana* and the *Os*U3 from *Oryza sativa*—conversely, the human-optimized coding sequence of h*Sp*Cas9 was modulated by the CaMV35S promoter. The mutation rate of the regenerated transgenic hairy roots was determined to be 50 percent based on the observations. A total of eight mutants, including two heterozygous mutants, two homozygous mutants, and five biallelic mutants, were present at the observed rate. The mutants derived from a cohort of 16 distinct transgenic hairy root lines, each induced by the sgRNA and regulated by the *At*U6 promoter. On the other hand, it was found that none of the 13 autonomous transgenic hairy root lines that were regulated by the sgRNA and under the control of the *Os*U3 promoter exhibited any mutant traits. Therefore, the proposal to identify appropriate synthetic or endogenous plant-specific promoters to regulate the expression of CRISPR/Cas9 has the potential to enhance gene editing efficacy. Additionally, the study of sixteen edited lines revealed a consequent decrease of ninety percent in the expression levels of the *Sm*RAS gene within the A8 homozygous mutant line, which is harbored a seven nucleotide deletion. Moreover, the other heterozygous and biallelic edited lines exhibited a reduction in expression levels ranging from around forty to eighty percent. Subsequently, metabolomics analysis showed a decrease in the phenolic acid content, specifically lithospermic acid B (LAB) and rosmarinic acid (RA), in homozygous mutants. When the potential target gene shows great similarity to other members of its gene family, the execution of functional analysis becomes more difficult. Consequently, investigating the of loss of function requires the knockout of multiple genes rather than just one. The *Sm*RAS gene, chosen as the potential target, is among the eleven members of the RAS gene family. The ten remaining RAS family genes continue to be involved in the transcription and translation, indicating that RA biosynthesis still takes place.

In their report published in 2022, Shiels et al. proposed implementing CRISPR/Cas system to target THC acid synthase gene (THCAS). This approach aims to create plants that are devoid of delta-9-tetrahydrocannabidiol (THC) but have high amounts of cannabidiol (CBD). The authors suggest that such plants could have increased commercial value, particularly in countries where there are strict regulations on THC content (Shiels et al., 2022) (Figure 2B). *Cannabis sativa* L., known as hemp, is a dioecious plant that has recently garnered significant attention due to its medical benefits, which have demonstrated economic viability. At present, the legalization of medical cannabis has been seen in over 50 nations, projecting a market worth of $20.2 billion within the timeframe of 2020-2025 (Aliekperova et al., 2020). The exploration of genome editing in *Cannabis sativa* remains an area that has not yet been thoroughly investigated, with the advent of improved the biosynthesis of cannabinoids and other specialized metabolites associated with various pathways. The increasing need for the creation and accumulation of cannabinoids presents an opportunity for the development of FDA-approved cannabidiol (CBD) and a potential reduction in the production of THC, which has been linked to various adverse consequences (Pertwee, 2008; Zuardi, 2008; Russo, 2011).

While omics analysis has identified multiple candidate genes within the cannabis production pathway that could potentially be targeted for genome editing, there is a scarcity of papers documenting successful and consistent transformation of cannabis tissues. Zhang et al. (2021) demonstrated an improvement in shoot regeneration efficiency by reprogramming of plant meristems, by overexpressing various genes, including Cannabis developmental regulator homologous to *Zm*WUS2, *Nb*STM, *Nb*IPT, *Os*GRF4, and *At*GIF, in immature embryo hypocotyls (Zhang et al., 2021). In order to develop albino-edited plants of Cannabis sativa, the CRISPR/Cas-mediated gene editing technique was employed, explicitly targeting the phytoene desaturase gene (*Cs*PDS1), commonly used as a marker gene. As a result, the *Cs*PDS1 gene knockout. According to Zhang et al. (2021), among the six sgRNA designs examined, sgRNA3, sgRNA4, and sgRNA5, which were specifically designed with Exon 6 of *Cs*PDS1, exhibited a greater editing frequency and efficiency compared to the other sgRNAs. These findings suggest the significance of the target site's position in influencing the editing outcomes (Zhang et al., 2021).

In light of this, the CRISPR/Cas9 tool is highly efficient for gene editing, especially in reverse genetics studies intended to elucidate gene function by generating of knockout lines. However, when using this system, it is important to consider a number of important factors carefully, such as choosing the target locus, making a suitable sgRNA, and selecting the right promoters to drive the sgRNA. Moreover, the investigation of gene functionality presents increased complexity when the potential target gene for knockout pertains toa gene family member.

3.2.2 Multiplex gene editing with CRISPR/Cas in medicinal plants

Multiplex editing of gene targets using CRISPR/Cas-based methods facilitates the investigation of genes that exhibit a high degree of redundancy. A more straightforward approach to studying the functional significance of a gene with numerous isoforms or gene family members in reverse genetics involves simultaneously targeting of several sgRNAs and Cas proteins inside a single expression vector system. When examining a gene family, it becomes crucial to investigate the functional role of an individual gene if it exhibits significant homology with other representatives of the gene family. Consequently, several genes should be knocked out to produce the desired phenotypic change; other gene family members could otherwise compensate. Zhou et al. (2021) conducted a study that employed CRISPR/Cas9 to operate a multiplex gene knockout to target the laccase gene family associated with *Salvia miltiorrhiza* by implementing a "dual-locus editing" strategy. *S. miltiorrhiza*, a plant commonly used for traditional

medication in China, is perceived for its biologically significant lipid-soluble compounds, namely tanshinones, as well as water-soluble phenolics such as rosmarinic acid (RA), salvianolic acid B (SAB), and salvianic acid (Danshensu). These compounds demonstrated efficacy in cardio-cerebral and vascular disease treatments (Zhou et al., 2021). Laccases, classified as members of the benzenediol oxygen reductases (EC 1.10.3.2), are part of the multicopper-oxidase gene family. They are closely associated with the oxidation and polymerization of monolignols, suggesting their potential involvement in the synthesis of SAB (Festa et al., 2008). As of now, a total of twenty-nine laccases, referred to as *Sm*LACs, have been obtained from the *S. miltiorrhiza* database. Among them, *Sm*LAC7 and *Sm*LAC20 have been specifically investigated to assess their impact on the production of SAB (Li et al., 2019). Nevertheless, the evaluation of plant phenotypes cannot be adequately influenced by the functional role of a single *Sm*LAC gene due to the prolixity of redundant laccase genes (Simões et al., 2020). The sgRNAs utilized in the dual-locus editing approach were specifically intended to target the Cu-oxidase_3 domain and Cu-oxidase_2 domain. Hence, the entire ORFs of twenty-nine *Sm*LACs were aligned using the BLAST homology algorithm (Li et al., 2019). It was interpreted that the laccase genes in *S. miltiorrhiza* were organized into seven distinct clusters. The sgRNA1 specifically targeted a gene belonging to group VII, whereas the sgRNA2 targeted genes from groups I, II, III, IV, and V. The subsequent examination of phenolic acid metabolism revealed that *Sm*LACs have a significant outcome in altering lignin production, which is essential for both root development and phenolic acid metabolism. A significant reduction was observed in the expression of key genes associated with phenolic acids synthesis, indicating the potential involvement of *Sm*LAC genes in this biosynthetic process (Zhou et al., 2021).

Nicotiana tabacum has two enzyme residues, namely core α (1,3)-fucose (FucT) and core β (1,2)-xylose (XylT), recognized as plant-derived glycoproteins, responsible for inducing an immunological or allergic response. The strategy is to eliminate the presence of FucT and XylT to provide a secure and therapeutic method for producing recombinant proteins using a plant-based expression platform. Previous research studies by Strasser et al. (2008) in *N. benthamiana* demonstrate the application of RNA interference to silence the FucT and XylT to achieve this purpose (Strasser et al., 2008). A TALEN-based strategy was also employed to induce knockout mutations in the *Nb*FucT and *Nb*XylT in Nicotiana benthamiana (Li et al., 2016). However, using of RNAi and TALEN methodologies could not result in a comprehensive depletion of both enzymes due to the partial concurrent inactivation of multiple genes. Hence, a focused multiplex CRISPR/Cas9 system was constructed to deactivate many genes simultaneously. Thus, a multiplex knockout strategy was implemented using CRISPR/Cas9, explicitly targeting the conserved region of four *FucT* and two *XylT* genes in BY-2 cells of *Nicotiana tabacum*, generating a knockout line encompassing a total of 12 alleles, including isoforms. In order to enhance the probability of inducing mutations across all isoforms, the sgRNAs were directed to edit the highly conserved domains of the targeted genes specifically. The sgRNA designs were chosen to aim for the first exon of the *Nt*XylT gene and the third exon of the *Nt*FucT gene. These sgRNAs were assembled for multiplexing, with a tRNA placed after the U6-driven sgRNA that is located upstream of the polycistron and pcoCas9 enzyme presided over by the promoter CaMV35S (Xie, Minkenberg, and Yang, 2015). The western blot analysis of the total cellular proteins, employing antibodies that specifically recognize FucT and XylT, indicated no detectable signal in the two lines (11,12). This observation suggests that the gene inactivation process was successful and complete. The MALDI-TOP analysis for examining the N-glycans in samples 11 and 12 did not exhibit any detectable levels of FucT and XylT residues, in contrast to the wild-type control line. The control line showed that over eighty three percent carried FucT residues and ninety one percent of XylT (Mercx et al., 2017).

In 2017, Li et al. employed CRISPR/Cas9 to induce deletion in the diterpene synthase gene (*Sm*CPS1) responsible for tanshinone synthesis in *S. miltiorrhiza* (Li et al., 2018). Tanshinones are a class of lipid-soluble chemicals that is found to possess properties that enhance blood circulation and have anti-inflammatory effects. Tanshinones utilize geranylgeranyl diphosphate (GGPP) as the shared precursor in taxol production. The cyclization of GGPP into tricyclic olefin miltiradiene has been demonstrated by two diterpene synthases, namely *Sm*CPS1 and *Sm*KSL1. This finding underscores the possibility of redirecting the biosynthetic pathway of GGPP towards the production of diterpenes, such as taxol by inhibiting the flux of GGPP towards tanshinone. Previously, RNAi gene silencing of *SmCPS1* showed a decrease in tanshinone production. Nevertheless, it remained observable in chimeric mutants, suggesting that a significant occurrence of non-specific gene silencing occurs due to sequence homology, hence facilitating the degradation of transcripts that are not the intended targets (Cheng et al., 2014; Cui et al., 2015). A recent study by McCarty et al. (2020) demonstrates using a multiplexed CRISPR/Cas vector system to simultaneously express multiple sgRNAs or Cas enzymes (McCarty et al., 2020). Three specific exon locations were chosen within the coding region (first, fourth, and eleventh) of the *Sm*CPS1 gene. Consequently, the three sgRNAs were developed to be controlled by the *At*U6-26 promoter and the expression cassette for *Sp*Cas9 by promoter CaMV35S. The proportion of independent transgenic T0 lines exhibiting successful mutations in *Sm*CPS1 was approximately 42.3 percent. Consequently, out of twenty-six independent transgenic hairy root lines, three lines were identified as homozygous mutants, while eight lines displayed chimeric mutations. Subsequent metabolomics analysis utilizing advanced mass spectrometry techniques unveiled the absence of specific tanshinones in the homozygous mutant lines, notably cryptotanshinone, tanshinone I, and tanshinone IIA. Therefore, the utilization of CRISPR/Cas9 technology to create knockouts effectively inhibits the flow of metabolic processes involving GGPP, thereby highlighting the potential for redirecting GGPP towards the production of essential diterpenes such as taxol (Li, Cui, et al., 2018).

The plant known as Comfrey, scientifically classified as *Symphytum officinale*, possesses medicinal attributes characterized by anti-inflammatory and analgesic effects. In addition to possessing pharmaceutically useful metabolites, pyrrolizidine alkaloids (PAs) in tissues render it poisonous to humans. The CRISPR/Cas9-mediated

knockout strategy successfully eliminated homospermidine synthase (*HSS*), which serves as the initial enzyme in the pathway of polyamine (PA) synthesis. The observed hairy root lines derived from the *hss* gene exhibited reduced homospermidine and PA levels upon the inactivation of only one *hss* allele. Furthermore, upon the inactivation of both alleles, no alkaloids were detectable. Three *At*U6-driven sgRNAs fabricated according to the target sites (exons three, seven, and eight) within the coding region of the *hss* gene were assembled along with *Sp*Cas9 driven by CaMV35S. The knockout lines showed no detectable PAs and 80% reduced levels of homospermidine, exhibiting that comfrey-derived economical phenolics such as rosmaniric acid and allantoin potentially be extracted from PA-free knockout lines without any lethal damage (Roeder et al., 2015; Zakaria et al., 2021). CRISPR/Cas-mediated engineering eliminates undesired multiple branching pathways by performing simultaneous gene knockouts by multiplexing (Alagoz et al., 2016; Sun et al., 2017). In the most recent study, six tomato genes were successfully deleted using 12 sgRNAs and a single CRISPR/Cas system expression cassette. Utilizing of an optimal technique to simultaneously target several target genes presents the potential for developing a highly efficient multiplex mutation system. For instance, the concurrent suppression of THCAS and CBDAs could be achieved to enhance the desired outcome (Čermák and Curtin, 2017; Deguchi et al., 2020).

3.3 Bottlenecks of using CRISPR/Cas for plant metabolic engineering

3.3.1 Off-target effects

CRISPR/Cas system allows precise and effective modification of DNA sequences. However, off-target effects in similar gene sequences can limit its application (Kleinstiver et al., 2016; Modrzejewski et al., 2020). A logistic regression model identified five factors influencing off-target occurrence: (1) the number of mismatches and sgRNA structure; (2) mismatch position; (3) GC-content; (4) nuclease variants; and (5) delivery methods. Increasing mismatches reduces off-target effects, and optimizing sgRNAs minimizes them (Modrzejewski et al., 2020). The optimal design of a sgRNA aims to enhance the efficacy of targeting the desired genomic site while limiting any possible off-target effects. The off-target occurrence rates exhibit a reduction of up to fifty-nine percent when a single mismatch is present, and this reduction becomes even more pronounced when four or more mismatches are present. The mismatches present in the seed sequence, namely within the first eight nucleotides near the PAM site, have been found to reduce the event of off-target effects considerably. The current understanding of the impact of GC content, nuclease variations such as Cas9 nickases, and delivery techniques on off-target effects is finite (Doench et al., 2016; Kimberland et al., 2018; Shen et al., 2019; Chen, 2019b).

3.3.2 The efficiency of CRISPR/Cas9 tool delivery

A significant obstacle to successful transformation is the effectiveness of CRISPR/Cas9 and sgRNA delivery to plant cells. The limited transformation efficiency and the challenge of selecting suitable plasmid vector systems hinder utilizing the CRISPR/Cas9 and sgRNA delivery system. There exist two strategies for surmounting the challenges. The initial approach involves Cas9 protein expression at the cellular level. Alternatively, a combination of Cas9-sgRNA can be prepared *in vitro* before the process of transformation or transfection, akin to the CRISPR ribonucleoprotein (RNP) system employed for protoplast delivery (as depicted in Figures 1B, H). An additional approach is the transportation of the sgRNA complexed with Cas9 protein to the nucleus through carrier nanoparticles (Dittmann et al., 2015; Tong et al., 2019).

3.3.3 CRISPR Cas system driving tool machinery

Improvements are needed in the driving machinery of the CRISPR/Cas tool, including promoters, terminators, selectable markers, and transformation protocols, for its effective utilization in endogenous plant systems. The present constraints in the intrinsic regulatory mechanisms inside medicinal plants pose obstacles to the synthesis of valuable bioactive compounds and inhibit the progress of CRISPR/Cas-mediated editing efficacy (Zhou et al., 2018; Jiang et al., 2017; Tong et al., 2019).

3.3.4 Linkage of phenotype to genotype

The gene editing CRISPR/Cas9 system provides both time and cost management advantages. However, linking the genotype to the phenotype poses a challenge in identifying the correct editing outcomes. To address this limitation, researchers have explored using CRISPR (MAGIC) systems and high-throughput biosensors for mapping genotypes and phenotypes (Lian et al., 2019).

3.3.5 Effective approaches to transformation and regeneration

There is a requirement for establishing efficacious plant transformation and regeneration protocols of enhancing the efficiency of gene editing tools like CRISPR/Cas9 in various plant species. The limitation of not having an effective tissue culture or transformation system in several plant species has resulted in notable hindrances to attaining a high transformation efficiency. In order to address this constraint, the utilization of *in vitro* tissue culture and protoplast regeneration approaches assumes a pivotal role in enhancing the transformation and regeneration competence of the CRISPR/Cas9 approach (Zhang et al., 2013; Altpeter et al., 2016; Mout et al., 2017).

4 Implications for improvement of transformation technologies for efficient delivery of gene-editing reagents in medicinal plants

The CRISPR/Cas gene editing system has revolutionized the understanding of plant genomes, enabling comprehensive profiling of medicinally valuable specialized metabolites involved in various plant secondary metabolic pathways. Despite the CRISPR/Cas

system's efficiency in editing the plant genome, the bottlenecks and challenges slow down further improvement and application (Cardi et al., 2023). The challenges are greater when implementing gene editing tools in novel or recalcitrant medicinal plants to establish a genotype-independent standard protocol for *in vitro* tissue culture, transformation, regeneration, and mutation detection in regenerants. The improvement in conventional *Agrobacterium*-mediated transformation depends on T-DNA delivery to overcome low transformation efficiency and *in vitro* regeneration ability but results in stable integration and gene expression leading to genetically modified plants (Nadakuduti and Enciso-Rodriguez, 2021). As a result, it reduces the regulatory mandates for acceptance as gene-edited plants until the elimination of the transgene by the backcrossing procedure. It also allows for establishing alternative transgene-free approaches to derive gene-edited plants with reduced off-targets or establishing genotype-independent in planta transformation protocols that do not rely on *in vitro* tissue culture and *de novo* regeneration. However, these methods require further work in medicinal plant species to advance gene-editing delivery (Han et al., 2021; Che et al., 2022).

Here, we discuss advances in establishing *in vitro* cultures, including hairy roots, tissue, callus, and cell suspension, to deliver gene editing reagents. We propose further using the information for CRISPR/Cas-mediated engineering of secondary metabolite pathways in medicinal plants.

4.1 *In vitro* differentiated-adventitious hairy roots culture and plant tissue culture

Agrobacterium rhizogenes is frequently used to induce hairy root development in medicinal plants. This technique facilitates the investigation of gene functionalities and the augmentation of yields of specialized metabolites (Tian, 2015; Hidalgo et al., 2018; Mi et al., 2020). The hairy root disease is induced by the infiltration of soil bacteria into wounded areas of dicot plants, leading to foreign gene integration into the plant's endogenous genome (Mi et al., 2020). Hairy root phenotypes are generated through T-DNA transfer in the root-inducing (Ri) plasmid, subsequent integration, and expression of a foreign gene from the Ri plasmid (Guillon et al., 2008). One of the most well-known effects of the gram-negative soil bacterium *Agrobacterium rhizogenes* is the development of hairy root disease. Hairy roots induced by *Agrobacterium rhizogenes* exhibit enhanced branching growth and differentiate from plagiotropic root development even without hormones, making them powerful tools in plant biotechnology (Srivastava and Srivastava, 2007). CRISPR/Cas9-mediated hairy root (HR) knockout lines of the homospermidine synthase (*HSS*) gene in *Symphytum officinale* were successfully generated by stable transformation using *Agrobacterium rhizogenes* (Zakaria et al., 2021). The resulting *hss* HR mutant line significantly reduced homospermidine and toxic pyrrolizidine alkaloids (PA) levels. CRISPR/Cas9 was also employed to generate hairy root mutant lines in *Atropa belladonna*, producing clinically significant tropane alkaloids, hyoscyamine, and scopolamine. For instance, an HR mutant line with a disrupted pyrrolidine ketide synthase (*PYKS*) gene, responsible for tropane skeleton construction, was developed to enhance generated pyrrolidine ketide synthesis (Hasebe et al., 2021).

In vitro plant tissue culture integrated with gene editing tool delivery, offers the distinct benefit of utilizing totipotent plant cells to derive chimeric regenerants and gene-edited plants, which can be employed to propagate rare, endangered plants (Figure 1E) (Rao and Ravishankar, 2002; Efferth, 2019; Cardi et al., 2023). Examples of CRISPR/Cas reagents delivery with *Agrobacterium*-mediated stable transformation via *in vitro* tissue culture methods are well established in a few medicinal plants, like *Atropa belladonna*, *Camelina sativa*, *Nicotiana tabacum* (Mercx et al., 2017; Jiang et al., 2017; Schachtsiek and Stehle, 2019; Zeng et al., 2021). Prospects to improve the transformation efficiency and *in vitro* regeneration to derive chimeric regenerants from edited cells for efficient gene editing in medicinal plants are still open to discussion (Cardi et al., 2023).

4.2 *In vitro* undifferentiated- callus and cell suspension culture

Plant-derived specialized metabolite production in microbial culture often suffers from erroneous post-translational modifications of proteins. Therefore, commercial plant suspension cell cultures have emerged as efficient and eco-friendly alternatives, successfully generating valuable metabolites. It is worth mentioning that there have been substantial advancements in genetic transformation and CRISPR/Cas-mediated genome editing in cell cultures of *Arabidopsis* and *Nicotiana tabacum*. Nevertheless, the idea of integrating gene editing tool delivery and cell cultures to manipulate heterologous biosynthetic pathways in medicinally-valued plants has yet to be widespread (Wu et al., 2021) (Figure 1D).

4.3 *Agrobacterium*-mediated transient transformation.

The cost-effective and time-efficient *Agrobacterium*-mediated transient transformation has gained interest for delivering gene editing reagents into plant systems because of the transient expression of Cas nucleases and deriving transgene-free gene-edited plant lines. Although polyethylene glycol (PEG)-mediated protoplast transformation for the transient delivery of CRISPR/Cas reagent is a rare example in medicinal plants. Few research studies show that there are initial attempts to establish *Agrobacterium*-mediated transient transformation in medicinal plants using reporter genes (Cardi et al., 2023). *Agrobacterium tumefaciens* injection and rapid *Agrobacterium*-mediated seedling transformation (FAST) were two novel techniques created by Xi and colleagues in 2020 to temporarily transform *Nicotiana benthamiana*, *Salvia miltiorrhiza*, and *Prunella vulgaris*. According to Xia et al. (2020), the expression of exogenous genes GUS and GFP was effectively achieved in *N. benthamiana* and *Prunella vulgaris*. However, the transient expression of exogenous genes in *S. miltiorrhiza* was hindered by a defense mechanism,

resulting in the unsuccessful transformation of the seedling system using *A. tumefaciens* (Xia et al., 2020).

5 Major transcription factors family regulating the metabolic pathways in medicinal plants- potential gene editing targets for enhancement of specialized metabolites

Plants synthesize plant-specialized metabolites, inclusive of flavonoids, nitrogen-containing alkaloids, terpenoids, phenolics, and polyphenolic compounds, in response to both biotic and environmental stressors (Chezem and Clay, 2016; Erb and Kliebenstein, 2020). These metabolites are synthesized via many metabolic pathways encompassing enzymatic reactions and gene regulatory mechanisms. Transcription factors have an indispensable regulatory role in the biosynthetic pathways and augmentation of bioactive metabolites in plants (Patra et al., 2013). Transcription factors selectively bind to specific cis-regulatory elements in promoter regions, which affects gene expression and makes it easier for transcriptional complexes, which are active in the transcription process to form (Yang et al., 2012). The comprehension of gene regulation of the metabolic biosynthesis networks and the functional roles of transcription factors leads to the potential discovery or enhancement of novel specialized metabolites in plants (Xu et al., 2016). Gene editing TF genes for generating knockout lines helps to dissect their molecular function in a particular metabolic pathway. TFs with adequate genomic information in the database, with known functional regulatory roles in a specific biosynthetic pathway, can exponentially activate the production of specialized metabolites by applying CRISPR/Cas gene editing tools. Here, we emphasize the major TFs family regulating critical secondary metabolic pathways in medicinal plants, which can provide prospective ideas for choice of potential gene editing targets to enhance the level of specialized metabolites in medicinal plants.

For instance, Luis et al., 2013 studied the involvement of APETALA2/ethylene-responsive element binding factors (AP2/ERF) and WRKY TFs in the regulation of artemisinin production, a highly potent herbal remedy for malaria (Luis et al., 2013). The AP2/ERF family encompasses four distinct sub-families, namely the AP2, ERF, RA, and dehydration-responsive element-binding protein (DREB) subfamilies. The regulation and response of plant metabolism to biotic and abiotic stress conditions are governed by AP2 and ERF TFs. The upregulation of ERF TFs that can connect with amorpha-4, 11-diene synthase (ADS) and CYP71AV1 motifs (P450 monooxygenase) resulted in elevated levels of artemisinin in genetically modified *Artemisia annua* plants (Yu et al., 2012).

The WRKY family of transcription factors comprises several genes that possess WRKY domains consisting of sixty highly conserved amino acids. These transcription factors are part of a substantial group that performs various functions in higher plants. They are involved in regulating plant signaling pathways, controlling plant secondary metabolism, and improving plant resistance to biotic and environmental factors (Jiang et al., 2017). The overexpression of the WRKY gene resulted in a 1.8-fold increase in artemisinin content in transgenic *A. annua* plants compared to the control. This increase in artemisinin production was attributed to the upregulation of the CYP71AV1 motif by the WRKY transcription factors (Rushton et al., 2010).

Another transcription factor family, basic helix-loop-helix (bHLH), demonstrates significant involvement in metabolic pathways associated with bioactive substances such as terpenoids, iridoids, and seco-iridoids. These compounds possess anti-cancerous, anti-microbial, and anti-inflammatory properties (Xu et al., 2016). The bHLH TFs also play notable functional roles in several biological processes, including plant growth, development, regulation of phytohormone levels, and maintenance of homeostasis.

The leucine zipper gene family, known as bZIP TFs, are important in plant growth, physiological processes, and stress responses. These TFs are distinguished by a conserved bZIP domain, consisting of two distinct structural features. The first feature is an essential region responsible for DNA binding, while the second feature is a leucine (Leu) zipper dimerization region (Zhang et al., 2015). The elucidation of the functions performed by these transcription factors can enhance the synthesis of plant-specialized metabolites.

The MYB TFs, on the other hand, are distinguished by the presence of a conserved DNA-binding helix-turn-helix domain, as well as a highly varied C-terminal activation domain. The architecture of these entities are classified into four distinct types: 1R (consisting of R1/2, R3-MYB, and MYB-related domains), 2R (comprising R2R3-MYB domains), 3R (composed of R1R2R3-MYB domains), and 4R (comprising R1R2R2R1/2-MYB domains). Transcription factors have essential and diverse functions in development of plant tissue and organs and their growth. These responsibilities encompass the production of xylem, guard cells, trichomes, and root hairs, as well as the synthesis of specialized metabolites, namely, alkaloids, terpenoids, flavonoids, and phenolics. A study conducted by Li et al. (2022) highlighted the significance of MYB TFs in *Camelia sinensis*. For instance, *Cs*MYB8 and 99 regulated flavonoid biosynthesis (including catechins, anthocyanins, and flavonol), *Cs*MYB85 and 86 controlled caffeine production, *Cs*MYB9 and 49 influenced theanine synthesis, *Cs*MYB110 managed carotenoid production, *Cs*MYB68, 147, 148, and 193 played roles in mono-/sesquiterpenoid volatiles production, *Cs*MYB164 and 192 were involved in lignin synthesis, and *Cs*MYB139, 162, and 198 impacted the synthesis of indolic compounds (Li, Xia, et al., 2022).

6 Future directions and perspective towards implementation of CRISPR/Cas system for the enhancement of specialized metabolites in medicinal plants

Medicinal plants are important in various fields, such as pharmacology, biomedicine, nutrition, and industry. Engineering metabolic circuits by eliminating competitive or

rate-limiting inhibitions to increase the production levels of plant-specialized metabolites are just a couple of the many possibilities for manipulating plant metabolic pathways. The CRISPR/Cas9 gene editing strategy shows promise in modifying target genes to meet commercial demands for enhanced metabolite production. Recent advancements in this technology have allowed for precise editing of the genome, epigenome, transcriptional regulation, and post-translational modifications. Advances in HDR for precision editing and chromosomal rearrangements driven by inversions or somatic cross-overs and base editing of single nucleotide have expanded the probability of adeptly using CRISPR/Cas technology.

Furthermore, CRISPR/Cas9 can generate large mutant libraries and enable gene knockout, knockdown (CRISPRi), and activation (CRISPRa). This technology can improve plant metabolic engineering by introducing DNA sequences from elite varieties, constructing new cell factories, and studying novel biosynthetic pathways. The ultimate objective is to increase the synthesis of pharmaceutically and commercially useful metabolites while decreasing the synthesis of harmful or undesirable compounds. Understanding the genetic, molecular, and biochemical processes is crucial for enhancing specialized metabolite accumulation in plant tissues. Various methods can be applied to enhance genome editing techniques, including using synthetic or endogenous promoters, terminators, and nuclear localization signals. Morineau et al. (2017) employed endogenous promoters from Camelina U3 and U6 to induce the production of Cas9 in a plant expression vector (Morineau et al., 2017). Therefore, the employment of CRISPR/Cas-mediated metabolic engineering approaches provide an opportunity to intensify the production of specialized metabolites in more significant quantities and unique configurations. Multiplex genome editing enables the concurrent manipulation of several genes, addressing the issue of excessive expression or suppression of individual genes or gene families. The emerging understanding of transcription factors and gene expression alterations holds promise for improvement in specialized metabolite production. The emerging understanding of transcription factors and gene expression alterations holds promise for improvement in specialized metabolite production.

The CRISPR/Cas system facilitates functional studies and enhances our understanding of novel pathways in medicinal plants, such as the cannabinoid and terpenoid pathways in *Cannabis sativa* (Bonini et al., 2018). Gracz-Bernaciak et al. (2021) propose applying the CRISPR/Cas methods to study latex production in laticiferous plants like *Chelidonium majus* L. and its antiviral activity (Gracz-Bernaciak et al., 2021). Establishment of efficient regeneration and stable transformation systems is of paramount importance in plant gene editing and the augmentation of specialized metabolites (Figures 1B–E). While the successful overexpression of endogenous genes using HDR remains a difficult task in medicinal plants, the utilization of CRISPR-based mutagenesis for endogenous gene modification has been widely established. Utilizing of cis-regulatory elements and untranslated regions (UTRs) presents various methodologies for precisely modulating gene expression. Recent studies have provided compelling evidence about the significant capabilities of manipulating cis-regulatory elements and untranslated regions (UTRs) in attaining meticulous regulation of gene expression. In brief, the CRISPR/Cas system is a gene editing technology that is both cost-effective and efficient, making it a promising candidate for further exploration in the field of medicinal plants (Shrestha et al., 2018; Wolter et al., 2019; Si et al., 2020).

7 Conclusion

The requirement for comprehensive genome and transcriptome data constrains the utilization of CRISPR/Cas9 gene editing. The acquisition of sequence data and the execution of functional investigations on metabolic pathways are paramount importance in the coordination of enzymatic activities and the regulation of branching metabolic pathways. Facilitating broader utilization of the CRISPR/Cas9 approaches in the context of therapeutic plants. This technological progress has the potential to facilitate the modification of essential genes responsible for some specific metabolite biosynthesis that have significant biological, nutritional, and pharmacological implications. The increasing understanding of genomes, transcriptomes, and proteomes associated with the production of the specialized metabolites, coupled with the progress made in gene editing technologies, presents an opportunity for the economically viable synthesis of these metabolites in medicinal plants of significant pharmaceutical importance. This potential development is approached within the context of regulatory and ethical frameworks.

Author contributions

J-YK: Conceptualization, Funding acquisition, Supervision, Writing – review & editing. SD: Conceptualization, Methodology, Writing – original draft, Writing – review & editing. MK: Conceptualization, Funding acquisition, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the National Research Foundation of Korea (Grant NRF 2022R1A2C3010331, 2021R1A5A8029490 and RS-2023-00301974), and the Technology Development Program (grant number, 20014582) funded by the Ministry of Trade, Industry & Energy (MOTIE, Korea).

Acknowledgments

We apologize to colleagues whose work could not be included owing to space constraints.

Conflict of interest

J-YK is a founder and CEO of Nulla Bio Inc.

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

References

Abudayyeh, O. O., Gootenberg, J., Essletzbichler, P., Han, S., Joung, J., Belanto, J., et al. (2017). RNA targeting with CRISPR–Cas13. *Nature.* 550, 280–284. doi: 10.1038/nature24049

Akama, K., Akter, N., Endo, H., Kanesaki, M., Endo, M., and Toki, S. (2020). An *in vivo* targeted deletion of the calmodulin-binding domain from rice glutamate decarboxylase 3 (OsGAD3) increases γ-aminobutyric acid content in grains. *Rice.* 13, 20. doi: 10.1186/s12284-020-00380-w

Alagoz, Y., Gurkok, T., Zhang, B., and Unver, T. (2016). Manipulating the biosynthesis of bioactive compound alkaloids for next-generation metabolic engineering in opium poppy using CRISPR-Cas 9 genome editing technology. *Sci. Rep.* 6, 30910. doi: 10.1038/srep30910

Aliekperova, N., Kosyachenko, K., and Kaniura, O. (2020). Perspectives on formation of medical cannabis market in Ukraine based on holistic approach. *J. Cannabis Res.* 2, 33. doi: 10.1186/s42238-020-00044-y

Allen, R. S., Miller, J., Chitty, J., Fist, A., Gerlach, L. W., and Philip, J. L. (2008). Metabolic engineering of morphinan alkaloids by over-expression and RNAi suppression of salutaridinol 7-O-acetyltransferase in opium poppy. *Plant Biotechnol. J.* 6, 22–30. doi: 10.1111/j.1467-7652.2007.00293.x

Altpeter, F., Springer, M. N., Bartley, E. L., Blechl, E. A., Brutnell, P. T., Citovsky, V., et al. (2016). Advancing crop transformation in the era of genome editing. *Plant Cell.* 28, 1510–1520. doi: 10.1105/tpc.16.00196

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

Bao, H., Chen, X., Lv, S., Jiang, P., Feng, J., Fan, P., et al. (2015). Virus-induced gene silencing reveals control of reactive oxygen species accumulation and salt tolerance in tomato by γ-aminobutyric acid metabolic pathway. *Plant Cell Environ.* 38, 600–613. doi: 10.1111/pce.12419

Barrangou, R., Birmingham, A., Wiemann, S., Beijersbergen, L. R., Hornung, V., and Smith, B. A. (2015). Advances in CRISPR-Cas9 genome engineering: lessons learned from RNA interference. *Nucleic Acids Res.* 43, 3407–3419. doi: 10.1093/nar/gkv226

Bhatnagar, N., Min, M., Choi, E., Kim, N., Moon, S. J., Yoon, I., et al. (2017). The protein phosphatase 2C clade A protein OsPP2C51 positively regulates seed germination by directly inactivating *Os*bZIP10. *Plant Mol. Biol.* 93, 389–401. doi: 10.1007/s11103-016-0568-2

Bhatta, P. B., and Malla, S. (2020). Improving horticultural crops via CRISPR/Cas9: current successes and prospects. *Plants (Basel).* 9, 1360. doi: 10.3390/plants9101360

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

Bonini, S. A., Premoli, M., Tambaro, S., Kumar, A., Maccarinelli, G., Memo, M., et al. (2018). Cannabis sativa: A comprehensive ethnopharmacological review of a medicinal plant with a long history. *J. Ethnopharmacol.* 227, 300–315. doi: 10.1016/j.jep.2018.09.004

Bourgaud, F., Gravot, A., Milesi, S., and Gontier, E. (2001). Production of plant secondary metabolites: a historical perspective. *Plant Science.* 161, 839–851. doi: 10.1016/S0168-9452(01)00490-3

Camerlengo, F., Frittelli, A., and Pagliarello, R. (2022). CRISPR towards a sustainable agriculture. *Encyclopedia.* 2 (1), 538–558. doi: 10.3390/encyclopedia2010036

Cardi, T., Murovec, J., Bakhsh, A., Boniecka, J., Bruegmann, T., Bull, S. E., et al. (2023). *Trends Plant Sci.* 28, 1144–1165. doi: 10.1016/j.tplants.2023.05.012

Čermák, T., and Curtin, S. J. (2017). Design and assembly of CRISPR/Cas9 reagents for gene knockout, targeted insertion, and replacement in wheat. *Methods Mol. Biol.* 1679, 187–212. doi: 10.1007/978-1-4939-7337-8_12

Chaurasiya, N. D., Sangwan, N. S., Sabir, F., Misra, L., and Sangwan, R. S. (2012). Withanolide biosynthesis recruits both mevalonate and DOXP pathways of isoprenogenesis in *Ashwagandha Withania somnifera* L. (Dunal). *Plant Cell Rep.* 31, 1889–1897. doi: 10.1007/s00299-012-1302-4

Chavez, A., Scheiman, J., Vora, S., Pruitt, B. W., Tuttle, M., Iyer, E., et al. (2015). Highly efficient Cas9-mediated transcriptional programming. *Nat. Methods* 12, 326–328. doi: 10.1038/nmeth.3312

Che, P., Wu, E., Simon, M. K., Anand, A., Lowe, K., Gao, H., et al. (2022). Wuschel2 enables highly efficient CRISPR/Cas-targeted genome editing during rapid *de novo* shoot regeneration in sorghum. *Commun. Biol.* 5, 344. doi: 10.1038/s42003-022-03308

Chen, S. (2019b). Minimizing off-target effects in CRISPR-Cas9 genome editing. *Cell Biol. Toxicology.* 35, 399–401. doi: 10.1007/s10565-019-09486-4

Chen, J., Qi, H., Li, J., Yi, Y., Chen, D., Hu, X., et al. (2014). Experimental study on *Dendrobium candidum* polysaccharides on promotion of hair growth. *Zhongguo Zhong Yao Za Zhi. China J. Chin. materia medica.* 39, 291–295.

Chen, K., Wang, Y., Zhang, R., Zhang, H., and Gao, C. (2019a). CRISPR/Cas genome editing and precision plant breeding in agriculture. *Annu. Rev. Plant Biol.* 70, 667–697. doi: 10.1146/annurev-arplant-050718-100049

Chen, Y., Zheng, Y. F., Lin, X. H., Zhang, J. P., Lin, F., and Shi, H. (2021). *Dendrobium* mixture attenuates renal damage in rats with diabetic nephropathy by inhibiting the PI3K/Akt/mTOR pathway. *Mol. Med. Rep.* 24, 590. doi: 10.3892/mmr.2021.12229

Cheng, Q., Su, P., Hu, Y., He, Y., Gao, W., and Huang, L. (2014). RNA interference-mediated repression of *Sm*CPS (copalyldiphosphate synthase) expression in hairy roots of *Salvia miltiorrhiza* causes a decrease of tanshinones and sheds light on the functional role of SmCPS. *Biotechnol. Lett.* 36, 363–369. doi: 10.1007/s10529-013-1358-4

Chezem, W., and Clay, K. N. (2016). Regulation of plant secondary metabolism and associated specialized cell development by MYBs and bHLHs. *Phytochemistry.* 131, 26–43. doi: 10.1016/j.phytochem.2016.08.006

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

Cui, G., Duan, L., Jin, B., Qian, J., Xue, Z., Shen, G., et al. (2015). Functional divergence of diterpene syntheses in the medicinal plant *salvia miltiorrhiza*. *Plant Physiol.* 169, 1607–1618. doi: 10.1104/pp.15.00695

Cui, B., Liang, Z., Liu, Y., Liu, F., and Zhu, J. (2012). Effects of ABA and its biosynthetic inhibitor fluridone on accumulation of penolic acids and activity of PAL and TAT in hairy root of *Salvia miltiorrhiza. Zhongguo Zhong Yao Za Zhi. China J. Chin. materia medica.* 37, 754–759.

Deguchi, M., Kane, S., Potlakayala, S., George, H., Proano, R., Sheri, V., et al. (2020). Metabolic engineering strategies of industrial hemp (Cannabis sativa L.): A brief review of the advances and challenges. *Front. Plant Science.* 11. doi: 10.3389/fpls.2020.580621

Deng, D., Yan, C., Pan, X., Mahfouz, M., Wang, J., Zhu, J., et al. (2012). Structural basis for sequence-specific recognition of DNA by TAL effectors. *Science.* 335, 720–723. doi: 10.1126/science.1215670

Dittmann, E., Gugger, M., Sivonen, K., and Fewer, D. P. (2015). Natural product biosynthetic diversity and comparative genomics of the cyanobacteria. *Trends Microbiol.* 23, 642–652. doi: 10.1016/j.tim.2015.07.008

Doench, J. G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E. W., Donovan, K. F., et al. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.* 34, 184–191. doi: 10.1038/nbt.3437

Dong, O. X., Yu, S., Jain, R., Zhang, N., Duong, P. Q., Butler, C., et al. (2020). Marker-free carotenoid-enriched rice generated through targeted gene insertion using CRISPR-Cas9. *Nat. Commun.* 11, 1178. doi: 10.1038/s41467-020-14981-y

Efferth, T. (2019). Biotechnology applications of plant callus cultures. *Engineering* 5, 50–59. doi: 10.1016/j.eng.2018.11.006

Erb, M., and Kliebenstein, D. J. (2020). Plant secondary metabolites as defenses, regulators, and primary metabolites: the blurred functional trichotomy. *Plant Physiol.* 184, 39–52. doi: 10.1104/pp.20.00433

Fairbairn, J. W., and Wassel, G. (1964). The alkaloids of Papaver somniferum L.—I.: Evidence for a rapid turnover of the major alkaloids. *Phytochemistry.* 3, 253–258. doi: 10.1016/S0031-9422(00)88047-4

Fauser, F., Schiml, S., and Puchta, H. (2014). Both CRISPR/Cas-based nucleases and nickases can be used efficiently for genome engineering in *Arabidopsis thaliana*. *Plant J.* 79, 348–359. doi: 10.1111/tpj.12554

Feng, S., Song, W., Fu, R., Zhang, H., Xu, A., and Li, J. (2018). Application of the CRISPR/Cas9 system in *Dioscorea zingiberensis*. *Plant Cell Tissue Organ Culture (PCTOC).* 135, 133–141. doi: 10.1007/s11240-018-1450-5

Festa, G., Autore, F., Fraternali, F., Giardina, P., and Sannia, G. (2008). Development of new laccases by directed evolution: functional and computational analyses. *Proteins.* 72, 25–34. doi: 10.1002/prot.21889

Gao, J., Jin, R., Wu, Y., Zhang, H., Zhang, D., Chang, Y., et al. (2002). Comparative study of tissue cultured Dendrobium protocorm with natural *Dendrobium candidum* on immunological function. *J. Chin.Med.Mater.* 25, 487–489.

Gao, S., Wang, B., Xie, S., Xu, X., Zhang, J., Pei, L., et al. (2020). A high-quality reference genome of wild *Cannabis sativa*. *Horticulture Res.* 7, 73. doi: 10.1038/s41438-020-0295-3

Gieger, P. L., and Hesse, H. (1833). Darstellung des atropins. *Ann. Chem. Pharm.* 5, 43–81. doi: 10.1002/jlac.18330050108

Globus, R., and Qimron, U. (2018). A technological and regulatory outlook on CRISPR crop editing. *J. Cell Biochem.* 119, 1291–1298. doi: 10.1002/jcb.26303

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. *Science.* 356, 438–442. doi: 10.1126/science.aam932

Gracz-Bernaciak, J., Mazur, O., and Nawrot, R. (2021). Functional studies of plant latex as a rich source of bioactive compounds: focus on proteins and alkaloids. *Int. J. Mol. Sci.* 22, 12427. doi: 10.3390/ijms222212427

Guillon, S., Trémouillaux-Guiller, J., Pati, P. K., and Gantet, P. (2008). "Hairy Roots: a Powerful Tool for Plant Biotechnological Advances," in *Bioactive Molecules and Medicinal Plants* (Springer Berlin Heidelberg, Berlin, Heidelberg).

Guo, M., Chen, H., Dong, S., Zhang, Z., and Luo, H. (2022). CRISPR-Cas gene editing technology and its application prospect in medicinal plants. *Chin. Med.* 17, 33. doi: 10.1186/s13020-022-00584-w

Guo, L., Winzer, T., Yang, X., Li, Y., Ning, Z., and He, Z. (2018). The opium poppy genome and morphinan production. *Science* 362, 343–347. doi: 10.1126/science.aat4096

Gurkok, T., Ozhuner, E., Parmaksiz, I., Özcan, S., Turktas, M., İpek, A., et al. (2016). Functional characterization of 4′OMT and 7OMT genes in BIA biosynthesis. *Front. Plant Science.* 7. doi: 10.3389/fpls.2016.00098

Han, Y., Broughton, S., Liu, L., Zhang, X., Zeng, J., He, X., et al. (2021). Highly efficient and genotype-independent barley gene editing based on anther culture. *Plant Commun.* 2, 100082. doi: 10.1016/j.xplc.2020.100082

Hasebe, F., Yuba, H., Hashimoto, T., Saito, K., Funa, N., and Shoji, T. (2021). CRISPR/Cas9-mediated disruption of the PYRROLIDINE KETIDE SYNTHASE gene reduces the accumulation of tropane alkaloids in *Atropa belladonna* hairy roots. *Biosci. Biotechnol. Biochem.* 85, 2404–2409. doi: 10.1093/bbb/zbab165

Hashimoto, T., and Yamada, Y. (1986). Hyoscyamine 6 beta-hydroxylase, a 2-oxoglutarate-dependent dioxygenase, in alkaloid-producing root cultures. *Plant Physiol.* 81, 619–625. doi: 10.1104/pp.81.2.619

Hassan, B. A. R. (2012). Medicinal plants (Importance and uses). *Pharm. Analytica Acta* 03, 10. doi: 10.4172/2153-2435.1000e139

Hayut, S. F., Bessudo, C. M., and Levy, A. A. (2017). Targeted recombination between homologous chromosomes for precise breeding in tomato. *Nat. Commun.* 8, 15605. doi: 10.1038/ncomms15605

Hidalgo, D., Sanchez, R., Lalaleo, L., Bonfill, M., Corchete, P., and Palazon, J. (2018). Biotechnological production of pharmaceuticals and biopharmaceuticals in plant cell and organ cultures. *CurrMedChem.* 25, 3577–3596. doi: 10.2174/0929867325666180309124317

Higashi, Y., and Saito, K. (2013). Network analysis for gene discovery in plant-specialized metabolism. *Plant Cell Environ.* 36, 1597–1606. doi: 10.1111/pce.12069

Huang, T. K., and Puchta, H. (2019). CRISPR/Cas-mediated gene targeting in plants: finally a turn for the better for homologous recombination. *Plant Cell Rep.* 38, 443–453. doi: 10.1007/s00299-019-02379-0

Hussein, R. A., and El-Anssary, A. A. (2018). "Plants Secondary Metabolites: The Key Drivers of the Pharmacological Actions of Medicinal Plants," in *Herbal Medicine*. Ed. F. Builders Philip (IntechOpen, Rijeka). doi: 10.5772/intechopen.76139

Iaffaldano, B., Yingxiao, Z., and Cornish, K. (2016). CRISPR/Cas9 genome editing of rubber producing dandelion *Taraxacum* kok-saghyz using *Agrobacterium rhizogenes* without selection. *Ind. Crops Products.* 89, 356–362. doi: 10.1016/j.indcrop.2016.05.029

Ilari, A., Franceschini, S., Bonamore, A., Arenghi, F. B., Botta, B. A., Macone, A., et al. (2009). Structural basis of enzymatic (S)-norcoclaurine biosynthesis. *J. Biol. Chem.* 284, 897–904. doi: 10.1074/jbc.M803738200

Jacobs, T. B., LaFayette, P. R., Schmitz, R. J., and Parrott, W. A. (2015). Targeted genome modifications in soybean with CRISPR/Cas9. *BMC Biotechnol.* 15, 16. doi: 10.1186/s12896-015-0131-2

Jadaun, J. S., Sangwan, N. S., Narnoliya, L. K., Singh, N., Bansal, S., Mishra, B., et al. (2017). Over-expression of DXS gene enhances terpenoidal secondary metabolite accumulation in rose-scented geranium and *Withania somnifera*: active involvement of plastid isoprenogenic pathway in their biosynthesis. *Physiol. Plant* 159, 381–400. doi: 10.1111/ppl.12507

Jäpelt, R. B., and Jakobsen, J. (2013). Vitamin D in plants: a review of occurrence, analysis, and biosynthesis. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2013.00136

Jiang, W. Z., Henry, I. M., Lynagh, P. G., Comai, L., Cahoon, E. B., and Weeks, D. P. (2017). Significant enhancement of fatty acid composition in seeds of the allohexaploid, Camelina sativa, using CRISPR/Cas9 gene editing. *Plant Biotechnol. J.* 15, 648–657. doi: 10.1111/pbi.12663

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M. F., Doudna, J., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science.* 337, 816–821. doi: 10.1126/science.1225829

Kadam, U. S., Shelake, R. M., Chavhan, R. L., and Suprasanna, P. (2018). Concerns regarding 'off-target' activity of genome editing endonucleases. *Plant Physiol. Biochem.* 131, 22–30. doi: 10.1016/j.plaphy.2018.03.027

Kajikawa, M., Shoji, T., Kato, A., and Hashimoto, T. (2011). Vacuole-localized berberine bridge enzyme-like proteins are required for a late step of nicotine biosynthesis in tobacco. *Plant Physiol.* 155, 2010–2022. doi: 10.1104/pp.110.170878

Kajikawa, M., Sierro, N., Kawaguchi, H., Bakaher, N., Ivanov, N. V., Hashimoto, T., et al. (2017). Genomic insights into the evolution of the nicotine biosynthesis pathway in tobacco. *Plant Physiol.* 174, 999–1011. doi: 10.1104/pp.17.00070

Kaushik, P., Gramazio, P., Vilanova, S., Raigón, M. D., Prohens, J., and Plazas, M. (2017). Phenolics content, fruit flesh colour and browning in cultivated eggplant, wild relatives and interspecific hybrids and implications for fruit quality breeding. *Food Res. Int.* 102, 392–401. doi: 10.1016/j.foodres.2017.09.028

Khusnutdinov, E., Sukhareva, A., Panfilova, M., and Mikhaylova, E. (2021). Anthocyanin biosynthesis genes as model genes for genome editing in plants. *Int. J. Mol. Sci.* 22, 8752. doi: 10.3390/ijms22168752

Kim, J. S. (2018). Precision genome engineering through adenine and cytosine base editing. *Nat. Plants.* 4, 148–151. doi: 10.1038/s41477-018-0115-z

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

Kim, J., and Kim, J. (2019). New era of precision plant breeding using genome editing. *Plant Biotechnol. Rep.* 13, 419–421. doi: 10.1007/s11816-019-00581-w

Kimberland, M. L., Hou, W., Alfonso-Pecchio, A., Wilson, S., Rao, Y., Zhang, S., et al. (2018). Strategies for controlling CRISPR/Cas9 off-target effects and biological variations in mammalian genome editing experiments. *J. Biotechnol.* 284, 91–101. doi: 10.1016/j.jbiotec.2018.08.007

Kleinstiver, B. P., Pattanayak, V., Prew, M. S., Tsai, S. Q., Nguyen, N. T., Zheng, Z., et al. (2016). High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature.* 529, 490–495. doi: 10.1038/nature16526

Knott, G. J., and Doudna, J. A. (2018). CRISPR-Cas guides the future of genetic engineering. *Science.* 361, 866–869. doi: 10.1126/science.aat5011

Koike, S., Matsukura, C., Takayama, M., Asamizu, E., and Ezura, H. (2013). Suppression of γ-aminobutyric acid (GABA) transaminases induces prominent GABA accumulation, dwarfism and infertility in the tomato (Solanum lycopersicum L.). *Plant Cell Physiol.* 54, 793–807. doi: 10.1093/pcp/pct035

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, 420–424. doi: 10.1038/nature17946

Konermann, S., Brigham, M. D., Trevino, A. E., Joung, J., Abudayyeh, O. O., and Barcena, C. (2015). Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature.* 517, 583–588. doi: 10.1038/nature14136

Kui, L., Chen, H., Zhang, W., He, S., Xiong, Z., Zhang, Y., et al. (2017). Building a genetic manipulation tool box for orchid biology: identification of constitutive promoters and application of CRISPR/Cas9 in the orchid, dendrobium officinale. *Front. Plant Science.* 7. doi: 10.3389/fpls.2016.02036

Kulkarni, R. (2016). Metabolic engineering: Biological art of producing useful chemicals. *Resonance* 21, 233–237. doi: 10.1007/s12045-016-0318-4

Kumar, R., Vashisth, D., Misra, A., Akhtar, M. Q., Jalil, S. U., Shanker, K., et al. (2016). RNAi down-regulation of cinnamate-4-hydroxylase increases artemisinin biosynthesis in Artemisia annua. *Sci. Rep.* 6, 26458. doi: 10.1038/srep26458

Kwon, M., Hodgins, C., Salama, E., Dias, K. R., Parikh, A., Ashlyn, V., et al. (2023). New insights into natural rubber biosynthesis from rubber-deficient lettuce mutants expressing goldenrod or guayule cis-prenyltransferase. *New Phytologist.* 239, 1098–1111. doi: 10.1111/nph.18994

Lan, X., Zeng, J., Liu, K., Zhang, F., Bai, G., Chen, M., et al. (2018). Comparison of two hyoscyamine 6 β-hydroxylases in engineering scopolamine biosynthesis in root cultures of Scopolia lurida. *Biochem. Biophys. Res. Commun.* 497, 25–31. doi: 10.1016/j.bbrc.2018.01.173

Li, B., Cui, G., Shen, G., Zhan, Z., Huang, L., Chen, J., et al. (2018). Targeted mutagenesis in the medicinal plant Salvia miltiorrhiza. *Scientific Reports* 7, 43320. doi: 10.1038/srep43320

Li, Q., Feng, J., Chen, L., Xu, Z., Zhu, Y., Wang, Y., et al. (2019). Genome-wide identification and characterization of *Salvia miltiorrhiza* laccases reveal potential targets for salvianolic acid b biosynthesis. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.00435

Li, T., Huang, S., Jiang, W. Z., Wright, D., Spalding, M. H., Weeks, D. P., et al. (2011). TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. *Nucleic Acids Res.* 39, 359–372. doi: 10.1093/nar/gkq704

Li, R., Li, R., Li, X., Fu, D., Zhu, B., Tian, H., et al. (2018). Multiplexed CRISPR/Cas9-mediated metabolic engineering of γ-aminobutyric acid levels in *Solanum lycopersicum*. *Plant Biotechnol. J.* 16, 415–427. doi: 10.1111/pbi.12781

Li, J., Scarano, A., Gonzalez, N. M., D'Orso, F., Yue, Y., Nemeth, K., et al. (2022). Biofortified tomatoes provide a new route to vitamin d sufficiency. *Nat. Plants*. 8, 611–616. doi: 10.1038/s41477-022-01154-6

Li, J., Stoddard, T. J., Demorest, Z. L., Lavoie, P., Luo, S., Clasen, B. M., et al. (2016). Multiplexed, targeted gene editing in *Nicotiana benthamiana* for glyco-engineering and monoclonal antibody production. *Plant Biotechnol. J.* 14, 533–542. doi: 10.1111/pbi.12403

Li, X., Wang, Y., Chen, S., Tian, H., Fu, D., Zhu, B., et al. (2018). Lycopene is enriched in tomato fruit by CRISPR/Cas9-mediated multiplex genome editing. *Front. Plant Science.* 9. doi: 10.3389/fpls.2018.00559

Li, F., Wang, W., Zhao, N., Xiao, B., Cao, P., Wu, X., et al. (2015). Regulation of nicotine biosynthesis by an endogenous target mimicry of MicroRNA in tobacco. *Plant Physiol.* 169, 1062–1071. doi: 10.1104/pp.15.00649

Li, P., Xia, E., Fu, J., Xu, Y., Zhao, X., Tong, W., et al. (2022). Diverse roles of MYB transcription factors in regulating secondary metabolite biosynthesis, shoot development, and stress responses in tea plants (*Camellia sinensis*). *Plant J.* 110, 1144–1165. doi: 10.1111/tpj.15729

Lian, J., Carl, S., Cao, M., HamediRad, M., and Zhao, H. (2019). Multi-functional genome-wide CRISPR system for high throughput genotype–phenotype mapping. *Nat. Commun.* 10, 5794. doi: 10.1038/s41467-019-13621-4

Liu, J., Gao, F., Ren, J., Lu, X., Ren, G., and Wang, R. (2017). A novel AP2/ERF transcription factor CR1 regulates the accumulation of vindoline and serpentine in *catharanthus roseus*. *Front. Plant Science.* 8. doi: 10.3389/fpls.2017.02082

Lossen, W., Lossen, W., Lossen, W., Lossen, W., and Ueber, (1864). 'das atropin'. *Ann. Chem. Pharm.* 131, 43–49. doi: 10.1002/jlac.18641310103

Mahas, A., Stewart, C. N. Jr., and Mahfouz, M. M. (2018). Harnessing CRISPR/Cas systems for programmable transcriptional and post-transcriptional regulation. *Biotechnol. Adv.* 36, 295–310. doi: 10.1016/j.bioteChadv.2017.11.008

Mahjoub, A., Hernould, M., Jounbes, J., Decendit, A., Mars, M., Barrieu, F., et al. (2009). Overexpression of a grapevine R2R3-MYB factor in tomato affects vegetative development, flower morphology and flavonoid and terpenoid metabolism. *Plant Physiol. Biochem.* 47, 551–561. doi: 10.1016/j.plaphy.2009.02.015

Mali, P., Luhan, Y., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., et al. (2013). RNA-guided human genome engineering via Cas9. *Science.* 339, 823–826. doi: 10.1126/science.1232033

Mani, M., Kandavelou, K., Dy, F. J., Durai, S., and Chandrasegaran, S. (2005). Design, engineering, and characterization of zinc finger nucleases. *Biochem. Biophys. Res. Commun.* 335, 447–457. doi: 10.1016/j.bbrc.2005.07.089

McCarty, N. S., Graham, A. E., Studená, 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

Mercx, S., Smargiasso, N., Chaumont, F., Pauw, E. D., Boutry, M., and Navarre, C. (2017). Inactivation of the β(1,2)-xylosyltransferase and the α(1,3)-fucosyltransferase genes in Nicotiana tabacum BY-2 Cells by a Multiplex CRISPR/Cas9 Strategy Results in Glycoproteins without Plant-Specific Glycans. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.00403

Mi, Y., Zhu, Z., Qian, G., Li, Y., Meng, X., and Xue, J. (2020). Inducing Hairy Roots by *Agrobacterium rhizogenes*-Mediated Transformation in Tartary Buckwheat (*Fagopyrum tataricum*). *J. Vis. Exp.* 11, (157). doi: 10.3791/60828

Modrzejewski, D., Hartung, F., Lehnert, H., Sprink, T., Kohl, C., and Keilwagen, J. (2020). Which factors affect the occurrence of off-target effects caused by the use of CRISPR/Cas: A systematic review in plants. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.574959

Mora-Vásquez, S., Wells-Abascal, G. G., Espinosa-Leal, C., Cardineau,, and García-Lara, S. (2022). Application of metabolic engineering to enhance the content of alkaloids in medicinal plants. *Metab. Eng. Commun.* 14, e00194. doi: 10.1016/j.mec.2022.e00194

Morineau, C., Bellec, Y., Tellier, F., Gissot, L., Kelemen, Z., and Nogué, F. (2017). Selective gene dosage by CRISPR-Cas9 genome editing in hexaploid *Camelina sativa*. *Plant Biotechnol.* 15, 729–739. doi: 10.1111/pbi.12671

Mout, R., Ray, M., Tonga, G., Lee, Y., Tay, T., Sasaki, K., et al. (2017). Direct cytosolic delivery of CRISPR/Cas9-ribonucleoprotein for efficient gene editing. *ACS Nano.* 11, 2452–2458. doi: 10.1021/acsnano.6b07600

Nadakuduti, S. S., and Enciso-Rodriguez, F. (2021). Advances in Genome Editing With CRISPR Systems and Transformation Technologies for Plant DNA Manipulation. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.637159

Niazian, M. (2019). Application of genetics and biotechnology for improving medicinal plants. *Planta.* 249, 953–973. doi: 10.1007/s00425-019-03099-1

Niu, Z., Zhu, F., Fan, Y., Li, C., Zhang, B., Zhu, S., et al. (2021). The chromosome-level reference genome assembly for *Dendrobium officinale* and its utility of functional genomics research and molecular breeding study. *Acta Pharm. Sin. B* 11, 2080–2092. doi: 10.1016/j.apsb.2021.01.019

Nonaka, S., Arai, C., Takayama, M., Matsukura, C., and Ezura, H. (2017). Efficient increase of ɣ-aminobutyric acid (GABA) content in tomato fruits by targeted mutagenesis. *Sci. Rep.* 7, 7057. doi: 10.1038/s41598-017-06400-y

Ossowski, S., Schwab, R., and Weigel, D. (2008). Gene silencing in plants using artificial microRNAs and other small RNAs. *Plant J.* 53, 674–690. doi: 10.1111/j.1365-313X.2007.03328.x

Paine, J. A., Shipton, C. A., Chaggar, S., Howells, R. M., Kennedy, M. J., and Vernon, G. (2005). Improving the nutritional value of Golden Rice through increased pro-vitamin A content. *Nat. Biotechnol.* 23, 482–487. doi: 10.1038/nbt1082

Patra, B., Schluttenhofer, C., Wu, Y., Pattanaik, S., and Yuan, L. (2013). Transcriptional regulation of secondary metabolite biosynthesis in plants. *Biochim. Biophys. Acta (BBA) - Gene Regul. Mechanisms.* 1829, 1236–1247. doi: 10.1016/j.bbagrm.2013.09.006

Paul, P., Singh, S. K., Patra, B., Sui, X., Pattanaik, S., and Yuan, L. (2017). A differentially regulated AP2/ERF transcription factor gene cluster acts downstream of a MAP kinase cascade to modulate terpenoid indole alkaloid biosynthesis in *Catharanthus roseus*. *New Phytol.* 213, 1107–1123. doi: 10.1111/nph.14252

Pertwee, R. G. (2008). The diverse CB1 and CB2 receptor pharmacology of three plant cannabinoids: delta9-tetrahydrocannabinol, cannabidiol and delta9-tetrahydrocannabivarin. *Br. J. Pharmacol.* 153, 199–215. doi: 10.1038/sj.bjp.0707442

Poupko, J. M., Baskin, S., and Moore, E. (2007). The pharmacological properties of anisodamine. *J. Appl. Toxicol.* 27, 116–121. doi: 10.1002/jat.1154

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

Pyne, M. E., Narcross, L., and Martin, V. J. J. (2019). Engineering plant secondary metabolism in microbial systems. *Plant Physiol.* 179, 844–861. doi: 10.1104/pp.18.01291

Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P., et al. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* 152, 1173–1183. doi: 10.1016/j.cell.2013.02.022

Raffan, S., Sparks, C., Huttly, A., Hyde, L., Martignago, D., Mead, A., et al. (2021). Wheat with greatly reduced accumulation of free asparagine in the grain, produced by CRISPR/Cas9 editing of asparagine synthetase gene *Ta*ASN2. *Plant Biotechnol. J.* 19, 1602–1613. doi: 10.1111/pbi.13573

Ramaswami, S. V., and Thangavelu, L. (2015). Pharmacological profile of tropane alkaloids. *J. Chem. Pharm. Res.* 7, 117–119.

Rao, S. R., and Ravishankar, G. A. (2002). Plant cell cultures: Chemical factories of secondary metabolites. *Biotechnol. Adv.* 20, 101–153. doi: 10.1016/s0734-9750(02)00007-1

Riaz, B., Chen, H., Wang, J., Du, L., Wang, K., and Ye, X. (2019). Overexpression of maize *Zm*C1 and *Zm*R transcription factors in wheat regulates anthocyanin biosynthesis in a tissue-specific manner. *Int. J. Mol. Sci.* 20, 5806. doi: 10.3390/ijms20225806

Roeder, E., Wiedenfeld, H., and Edgar, J. A. (2015). Pyrrolizidine alkaloids in medicinal plants from North America. *Die Pharm.* 70, 357–367. doi: 10.1691/ph.2015.4873

Rushton, P. J., Somssich, I. E., Ringler, P., and Shen, Q. J. (2010). WRKY transcription factors. *Trends Plant Sci.* 15, 247–258. doi: 10.1016/j.tplants.2010.02.006

Russo, E. B. (2011). Taming THC: potential cannabis synergy and phyto cannabinoid-terpenoid entourage effects. *Br. J. Pharmacol.* 163, 1344–1364. doi: 10.1111/j.1476-5381.2011.01238.x

Samad, A. F. A., Sajad, M., Nazaruddin, N., Fauzi, I. A., Murad, A. M. A., Zainal, Z., et al. (2017). MicroRNA and transcription factor: key players in plant regulatory network. *Front. Plant Science.* 8. doi: 10.3389/fpls.2017.00565

Sanchez, S., and Demain, A. L. (2008). Metabolic regulation and overproduction of primary metabolites. *Microb. Biotechnol.* 1, 283–319. doi: 10.1111/j.1751-7915.2007.00015.x

Sangwan, R. S., Chaurasiya, N. D., Lal, P., Misra, L., Tuli, R., and Sangwan, N. S. (2008). Withanolide A is inherently *de novo* biosynthesized in roots of the medicinal plant Ashwagandha (Withania somnifera) 133, 278–287. doi: 10.1111/j.1399-3054.2008.01076.x

Sangwan, R. S., Chaurasiya, N. D., Lal, P., Misra, L., Uniyal, G. C., Tuli, R., et al. (2007). Withanolide A biogeneration in *in vitro* shoot cultures of ashwagandha (*Withania somnifera* DUNAL), a main medicinal plant in Ayurveda. *Chem. Pharm. Bull.* 55, 1371–1375. doi: 10.1248/cpb.55.1371

Sangwan, R., Chaurasiya, N. D., Misra, L., Lal, P., Uniyal, G. C., Sharma, R., et al. (2004). Phytochemical variability in commercial herbal products and preparations of *Withania somnifera* (Ashwagandha). *Curr. Sci.* 86, 461–465.

Sangwan, N. S., Jadaun, J. S., Tripathi, S., Mishra, B., Narnoliya, L. K., and Sangwan, R. S. (2018). Plant metabolic engineering. *Omics Technol. Bio-Engineering.*, 143–175. doi: 10.1016/B978-0-12-815870-8.00009-7

Sangwan, N. S., Mishra, L. N., Tripathi, S., and Kushwaha, A. K. (2014a). Omics of secondary metabolic pathways in *withania somnifera* dunal (Ashwagandha). *Omics Appl. Crop Science.* doi: 10.1201/b16352-12

Sangwan, N. S., and Sangwan, R. S. (2014b). Secondary metabolites of traditional medical plants: A case study of ashwagandha (*Withania somnifera*). *Appl. Plant Cell Biol.*, 325–367. doi: 10.1007/978-3-642-41787-0_11

Sashidhar, N., Harloff, H. J., Potgieter, L., and Jung, C. (2020). Gene editing of three BnITPK genes in tetraploid oilseed rape leads to significant reduction of phytic acid in seeds. *Plant Biotechnol. J.* 18, 2241–2250. doi: 10.1111/pbi.13380

Schachtsiek, J., and Stehle, F. (2019). Nicotine-free, nontransgenic tobacco (*Nicotiana tabacum* l.) edited by CRISPR-Cas9. *Plant Biotechnol. J.* 17, 2228–2230. doi: 10.1111/pbi.13193

Sharafi, A., Sohi, H. H., Mousavi, A., Azadi, P., Khalifani, B. H., and Razavi, K. (2013). Metabolic engineering of morphinan alkaloids by over-expression of codeinone reductase in transgenic hairy roots of *Papaver bracteatum*, the Iranian poppy. *Biotechnol. Lett.* 35, 445–453. doi: 10.1007/s10529-012-1080-7

Shen, C. C., Hsu, M., Chang, C., Lin, M., Hwu, J., Tu, Y., et al. (2019). Synthetic switch to minimize CRISPR off-target effects by self-restricting Cas9 transcription and translation. *Nucleic Acids Res.* 47, e13–e13. doi: 10.1093/nar/gky1165

Shi, M., Du, Z., Hua, Q., and Kai, G. (2021). CRISPR/Cas9-mediated targeted mutagenesis of bZIP2 in *Salvia miltiorrhiza* leads to promoted phenolic acid biosynthesis. *Ind. Crops Products.* 167, 113560. doi: 10.1016/j.indcrop.2021.113560

Shiels, D., Prestwich, B., Koo, O., Kanchiswamy, C. N., O'Halloran, R., and Badmi, R. (2022). Hemp genome editing—Challenges and opportunities. *Front. Genome Editing* 4. doi: 10.3389/fgeed.2022.823486

Shlush, B. I., Samach, A., Melamed-Bessudo, C., Ben-Tov, D., Dahan-Meir, T., Filler-Hayut, S., et al. (2021). CRISPR/Cas9 induced somatic recombination at the CRTISO locus in tomato. *Genes (Basel).* 12, 59. doi: 10.3390/genes12010059

Shrestha, A., Khan, A., and Dey, N. (2018). cis-trans engineering: advances and perspectives on customized transcriptional regulation in plants. *Mol. Plant* 11, 886–898. doi: 10.1016/j.molp.2018.05.008

Si, X., Zhang, H., Wang, Y., Chen, K., and Gao, C. (2020). Manipulating gene translation in plants by CRISPR-Cas9-mediated genome editing of upstream open reading frames. *Nat. Protoc.* 15, 338–363. doi: 10.1038/s41596-019-0238-3

Simões, M. S., Carvalho, G. G., Ferreira, S. S., Hernandes-Lopes, J., de Setta, N., and Cesarino, I. (2020). Genome-wide characterization of the laccase gene family in Setaria viridis reveals members potentially involved in lignification. *Planta.* 251, 46. doi: 10.1007/s00425-020-03337-x

Singer, A. C., Crowley, D. E., and Thompson, I. P. (2003). Secondary plant metabolites in phytoremediation and biotransformation. *Trends Biotechnol.* 21, 123–130. doi: 10.1016/S0167-7799(02)00041-0

Singh, S. K., Patra, B., Paul, P., Liu, Y., Pattanaik, S., and Yuan, L. (2020). Revisiting the ORCA gene cluster that regulates terpenoid indole alkaloid biosynthesis in *Catharanthus roseus*. *Plant Sci.* 293, 110408. doi: 10.1016/j.plantsci.2020.110408

Singh, P., Prasad, R., Tewari, R., Jaidi, M., Kumar, S., Rahman, L., et al. (2018). Silencing of quinolinic acid phosphoribosyl transferase (QPT) gene for enhanced production of scopolamine in hairy root culture of *Duboisia leichhardtii*. *Sci. Rep.* 8, 13939. doi: 10.1038/s41598-018-32396-0

Sinha, S., Sandhu, K., Bisht, N., Nailwal, T., Saini, I., and Kaushik, P. (2019). Ascertaining the paradigm of secondary metabolism enhancement through gene level modification in therapeutic plants. *J. Young Pharmacists.* 11, 337–343. doi: 10.5530/jyp.2019.11.70

Snowden, C. J., Thomas, B., Baxter, C. J., Smith, A. C., and Sweetlove, L. J. (2013). A tonoplast Glu/Asp/GABA exchanger that affects tomato fruit amino acid composition. *Plant J* 81, 651–660. doi: 10.1111/tpj.12766

Sonawane, P. D., Pollier, J., Panda, S., Szymanski, J., Massalha, H., Yona, M., et al. (2016). Plant cholesterol biosynthetic pathway overlaps with phytosterol metabolism. *Nat. Plants.* 3, 16205. doi: 10.1038/nplants.2016.205

Sparvoli, F., and Cominelli, E. (2015). Seed biofortification and phytic acid reduction: A conflict of interest for the plant? *Plants.* 4, 728–755. doi: 10.3390/plants4040728

Springob, K., and Kutchan, T. M. (2009). Introduction to the different classes of natural products. *Plant derived Natural Prod.* doi: 10.1007/978-0-387-85498-4_1

Srivastava, S., and Srivastava, A. K. (2007). Hairy root culture for mass-production of high-value secondary metabolites. *Crit. Rev. Biotechnol.* 27, 29–43. doi: 10.1080/07388550601173918

Strasser, R., Stadlmann, J., Schähs, M., Stiegler, G., Quendler, H., Mach, L., et al. (2008). Generation of glyco-engineered *Nicotiana benthamiana* for the production of monoclonal antibodies with a homogeneous human-like N-glycan structure. *Plant Biotechnol. J.* 6, 392–402. doi: 10.1111/j.1467-7652.2008.00330.x

Strickler, S. R., Bombarely, A., and Mueller, L. A. (2012). Designing a transcriptome next-generation sequencing project for a non model plant species. *Am. J. Bot.* 99, 257–266. doi: 10.3732/ajb.1100292

Sun, Y., Guiai, J., Liu, Z., Zhang, X., Li, J., Guo, X., et al. (2017). Generation of high-amylose rice through CRISPR/Cas9-mediated targeted mutagenesis of starch branching enzymes. *Front. Plant Science.* 8. doi: 10.3389/fpls.2017.00298

Tian, L. (2015). Using hairy roots for production of valuable plant secondary metabolites. *Adv. Biochem. Eng. Biotechnol.* 149, 275–324. doi: 10.1007/10_2014_298

Tong, Y., Weber, T., and Lee, S. Y. (2019). CRISPR/Cas-based genome engineering in natural product discovery. *Natural Product Rep.* 36, 1262–1280. doi: 10.1039/c8np00089a

Trethewey, R. N. (2004). Metabolite profiling as an aid to metabolic engineering in plants. *Curr. Opin. Plant Biol.* 7, 196–201. doi: 10.1016/j.pbi.2003.12.003

UN Report (2022) Global hunger numbers rose to as many as 828 million in 2021. Available online at: https://www.who.int/news/item/06-07-2022-un-report–global-hunger-numbers-rose-to-as-many-as-828-million-in-2021.

Valva, V., Sabato, S., and Gigliano, G. S. (1985). Morphology and alkaloid chemistry of *Papaver setigerum* DC. (Papaveraceae). *Taxon.* 34, 191–196. doi: 10.2307/1221777

Verpoorte, R., Contin, A., and Memelink, J. (2002). Biotechnology for the production of plant secondary metabolites. *Phytochem. Rev.* 1, 13–25. doi: 10.1023/A:1015871916833

Wang, L., Lu, W., Ran, L., Dou, L., Yao, S., Hu, J., et al. (2019). R2R3-MYB transcription factor MYB6 promotes anthocyanin and proanthocyanidin biosynthesis but inhibits secondary cell wall formation in *Populus tomentosa*. *Plant J.* 99, 733–751. doi: 10.1111/tpj.14364

Wilson, S. A., and Roberts, S. C. (2014). Metabolic engineering approaches for production of biochemicals in food and medicinal plants. *Curr. Opin. Biotech.* 26, 174–182. doi: 10.1016/j.copbio.2014.01.006

Wolter, F., Schindele, P., and Puchta, H. (2019). Plant breeding at the speed of light: the power of CRISPR/Cas to generate directed genetic diversity at multiple sites. *BMC Plant Biol.* 19, 176. doi: 10.1186/s12870-019-1775-1

Wright, D. A., Townsend, A. J., Winfrey, R. J., Irwin, P. A., Rajagopal, J., Lonosky, P. M., et al. (2004). High-frequency homologous recombination in plants mediated by zinc-finger nucleases. *Plant J* 44, 693–705. doi: 10.1111/j.1365-313X.2005.02551.x

Wu, T., Kerbler, S. M., Fernie, A. R., and Zhang, Y. (2021). Plant cell cultures as heterologous bio-factories for secondary metabolite production. *Plant Commun.* 2, 100235. doi: 10.1016/j.xplc.2021.100235

Xia, P., Hu, W., Liang, T., Yang, D., and Liang, Z. (2020). An attempt to establish an *Agrobacterium*-mediated transient expression system in medicinal plants. *Protoplasma.* 257, 1497–1505. doi: 10.1007/s00709-020-01524-x

Xie, K., Minkenberg, B., and Yang, Y. (2015). Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. *Proc. Natl. Acad. Sci. U.S.A.* 112, 3570–3575. doi: 10.1073/pnas.1420294112

Xu, H., Song, J., Luo, H., Zhang, Y., Li, Q., Zhu, Y., et al. (2016). Analysis of the genome sequence of the medicinal plant *salvia miltiorrhiza*. *Mol. Plant* 9, 949–952. doi: 10.1016/j.molp.2016.03.010

Yang, C., Fang, X., Wu, X., Mao, Y., Wang, L., and Chen, X. (2012). Transcriptional regulation of plant secondary metabolism. *J. Integr. Plant Biol.* 54, 703–712. doi: 10.1111/j.1744-7909.2012.01161.x

Yang, L., Wen, K., Ruan, X., Zhao, Y., Wei, F., and Wang, Q. (2018). Response of plant secondary metabolites to environmental factors. *Molecules* 23, 762. doi: 10.3390/molecules23040762

Yu, J., Tu., L., Subburaj, S., Bae, S., and Lee, G. (2012). Simultaneous targeting of duplicated genes in *Petunia* protoplasts for flower color modification via CRISPR-Cas9 ribonucleoproteins. *MolPlant.* 40, 1037–1045. doi: 10.1007/s00299-020-02593-1

Yuan, H., Ma, Q., Ye, L., and Piao, G. (2016). The traditional medicine and modern medicine from natural products. *Molecules.* 21, 559. doi: 10.3390/molecules21050559

Yun, D. J., and Yamada, Y. (1992). Metabolic engineering of medicinal plants: transgenic *Atropa belladonna* with an improved alkaloid composition. *Proc. Natl. Acad. Sci. U.S.A.* 89, 11799–11803. doi: 10.1073/pnas.89.24.11799

Zakaria, M. M., Schemmerling, B., and Ober, D. (2021). CRISPR/Cas9-mediated genome editing in comfrey (*Symphytum officinale*) hairy roots results in the complete eradication of pyrrolizidine alkaloids. *Molecules.* 26, 1498. doi: 10.3390/molecules26061498

Zeng, L., Zhang, Q., Jiang, C., Zheng, Y., Zuo, Y., Qin, J. Z., et al. (2021). Development of *Atropa belladonna* L. Plants with High-Yield Hyoscyamine and without Its Derivatives Using the CRISPR/Cas9 System. *Int. J. Mol. Sci.* 22, 1731. doi: 10.3390/ijms22041731

Zhang, F., Fu, X., Lv, Z., Lu, X., Shen, Q., Zhang, L., et al. (2015). A basic leucine zipper transcription factor, AabZIP1, connects abscisic acid signaling with artemisinin biosynthesis in. *Artemisia annua*. *Mol. Plant* 8, 163–175. doi: 10.1016/j.molp.2014.12.004

Zhang, L., Jing, F., Li., F., Li, M., Wang, Y., Wang, G., et al. (2009). Development of transgenic *Artemisia annua* (Chinese wormwood) plants with an enhanced content of artemisinin, an effective anti-malarial drug, by hairpin-RNA-mediated gene silencing. *Biotechnol. Appl. Biochem.* 52, 199–207. doi: 10.1042/BA20080068

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, 778–794. doi: 10.1038/s41477-019-0461-5

Zhang, H., Si, X., Ji, X., Fan, R., Liu, J., Chen, K., et al. (2018b). Genome editing of upstream open reading frames enables translational control in plants. *Nat. Biotechnol.* 36, 894–898. doi: 10.1038/nbt.4202

Zhang, F., Xiang, L., Yu, Q., Zhang, H., Zhang, T., Zeng, J., Xu, Y., et al. (2018a). ARTEMISININ BIOSYNTHESIS PROMOTING KINASE 1 positively regulates artemisinin biosynthesis through phosphorylating *Aa*bZIP1. *J. Exp. Bot.* 69, 1109–1123. doi: 10.1093/jxb/erx444

Zhang, X., Xu, G., Cheng, C., Lei, L., Sun, J., et al. (2021). Establishment of an *Agrobacterium*-mediated genetic transformation and CRISPR/Cas9-mediated targeted mutagenesis in Hemp (Cannabis Sativa L.). *Plant Biotechnol. J.* 19, 1979–1987. doi: 10.1111/pbi.13611

Zhang, Y., Zhang, F., Li, X., Baller, J., Qi, Y., Starker, C., et al. (2013). Transcription activator-like effector nucleases enable efficient plant genome engineering. *Plant Physiol.* 161, 20–27. doi: 10.1104/pp.112.205179

Zhou, Z., Li, Q., Xiao, L., Wang, Y., Feng, J., Bu, Q., et al. (2021). Multiplexed CRISPR/Cas9-mediated knockout of laccase genes in *salvia miltiorrhiza* revealed their roles in growth, development, and metabolism. *Front. Plant Science.* 12. doi: 10.3389/fpls.2021.647768

Zhou, Z., Tan, H., Li, Q., Chen, J., Gao, S., Wang, Y., et al. (2018). CRISPR/Cas9-mediated efficient targeted mutagenesis of RAS in *Salvia miltiorrhiza*. *Phytochemistry.* 148, 63–70. doi: 10.1016/j.phytochem.2018.01.015

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

Zuardi, A. W. (2008). Cannabidiol: from an inactive cannabinoid to a drug with wide spectrum of action. *Br. J. Psychiatry* 30, 271–280. doi: 10.1590/s1516-44462008000300015

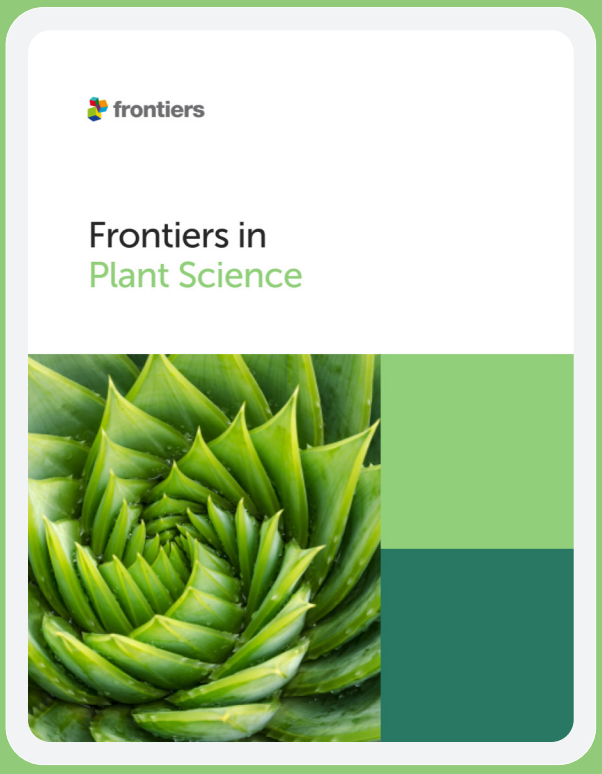
Frontiers in
Plant Science

Cultivates the science of plant biology and its applications

The most cited plant science journal, which advances our understanding of plant biology for sustainable food security, functional ecosystems and human health.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact